PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12Q 1/68, 1/70, C12N 7/00, 7/01, 7/02, G01N 33/564, C07K 1/00

A1

(11) International Publication Number: **WO 98/59074**

(43) International Publication Date: 30 December 1998 (30.12.98)

(21) International Application Number: PCT/US98/12990

(22) International Filing Date: 23 June 1998 (23.06.98)

(30) Priority Data:
60/050,548 23 June 1997 (23.06.97) US
60/057,606 4 September 1997 (04.09.97) US

(71) Applicant: EMORY UNIVERSITY [US/US]; 2009 Ridgewood Drive, Atlanta, GA 30322 (US).

(72) Inventors: NOVEMBRE, Francis, J.; 195 Chateau Way, Smyrna, GA 30082 (US). MCCLURE, Harold, M.; 2301 Poplar Springs Drive, Atlanta, GA 30319 (US).

(74) Agents: FERBER, Donna, M. et al.; Greenlee, Winner and Sullivan, P.C., Suite 201, 5370 Manhattan Circle, Boulder, CO 80303 (US).

(81) Designated States: AU, CA, CN, JP, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.

(54) Title: HUMAN IMMUNODEFICIENCY VIRUSES CAUSING AIDS IN A NONHUMAN PRIMATE

(57) Abstract

An isolated human immunodeficiency virus (HIV) type 1, having the identifying characteristics of HIV isolate JC and assigned AIDS Reagent Program Catalog Number 3523, was isolated from an HIV-infected chimpanzee that developed AIDS. This chimpanzee represents the first known animal model of HIV-1 induced AIDS. The substantially full-length (infectious) nucleotide sequences of the HIV-1_{JC} and of HIV-1_{NC} molecular clones are provided herein. This HIV-1_{JC} and HIV-1_{NC} isolates are useful for the preparation of recombinant, attenuated and subunit vaccines, as well as for the preparation of challenge stocks. It is also used as a diagnostic reagent in screening for the presence of HIV-1 in biological samples.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania
AM	Armenia
AT	Austria
AU	Australia
AZ	Azerbaijan
BA	Bosnia and Herzegovina
BB	Barbados
BE	Belgium
BF	Burkina Faso
BG	Bulgaria
BJ	Benin
BR	Brazil
BY	Belarus
CA	Canada
CF	Central African Republic
CG	Congo
CH	Switzerland
CI	Côte d'Ivoire
CM	Cameroon
CN	China
CU	Cuba
CZ	Czech Republic
DE	Germany
DK	Denmark
EE	Estonia
ES	Spain
FI	Finland
FR	France
GA	Gabon
GB	United Kingdom
GE	Georgia
GH	Ghana
GN	Guinea
GR	Greece
HU	Hungary
IE	Ireland
IL	Israel
IS	Iceland
IT	Italy
JP	Japan
KE	Kenya
KG	Kyrgyzstan
KP	Democratic People's Republic of Korea
KR	Republic of Korea
KZ	Kazakstan
LC	Saint Lucia
LI	Liechtenstein
LK	Sri Lanka
LR	Liberia
LS	Lesotho
LT	Lithuania
LU	Luxembourg
LV	Latvia
MC	Monaco
MD	Republic of Moldova
MG	Madagascar
MK	The former Yugoslav Republic of Macedonia
ML	Mali
MN	Mongolia
MR	Mauritania
MW	Malawi
MX	Mexico
NE	Niger
NL	Netherlands
NO	Norway
NZ	New Zealand
PL	Poland
PT	Portugal
RO	Romania
RU	Russian Federation
SD	Sudan
SE	Sweden
SG	Singapore
SI	Slovenia
SK	Slovakia
SN	Senegal
SZ	Swaziland
TD	Chad
TG	Togo
TJ	Tajikistan
TM	Turkmenistan
TR	Turkey
TT	Trinidad and Tobago
UA	Ukraine
UG	Uganda
US	United States of America
UZ	Uzbekistan
VN	Viet Nam
YU	Yugoslavia
ZW	Zimbabwe

HUMAN IMMUNODEFICIENCY VIRUSES CAUSING AIDS IN A NONHUMAN PRIMATE

CROSS REFERENCE TO RELATED APPLICATIONS

This application claims priority from United States Provisional Patent Application No. 60/050,548, filed June 23, 1997 and from United States Provisional Patent Application No. 60/057,606, filed September 4, 1997.

ACKNOWLEDGEMENT OF FEDERAL RESEARCH SUPPORT

This invention was made, at least in part, with funding from Public Health Service Grant RR-00165 from the National Center for Research Resources. The United States government may have certain rights in this invention.

BACKGROUND OF THE INVENTION

This invention concerns unique isolates of human immunodeficiency virus type 1 ($HIV\text{-}1_{JC}$) which are highly infectious *in vivo* and produce acquired immune deficiency syndrome (AIDS) in a nonhuman primate.

The human immunodeficiency viruses types 1 and 2 (HIV-1, HIV-2) are retroviruses which have been implicated as the causative agents of acquired immune deficiency syndrome (AIDS) (Barre-Sinoussi et al. [1983] Science **220**:868-871). While skepticism about the exact cause of AIDS has arisen (Duesberg [1991] Proc. Natl. Acad. Sci. **88**:1575-1579), a large amount of data has accumulated which supports HIV as the biologic agent of this disease. The most convincing evidence includes: mother to child transmission; transmission

via blood transfusion; and transmission via contaminated blood products (Curran et al. [1984] N. Engl. J. Med. **310**:69-75). The development of disease in an animal inoculated with HIV-1 would provide confirmatory evidence for the etiology of AIDS. While AIDS-like disease has been recently demonstrated in HIV-2-infected baboons (Barnett et al. [1994] Science **266**:642-646), to date, no species, other than humans, has developed AIDS following infection with HIV-1. Thus, the most appropriate animal model for HIV-1 infection currently does not exist. However, an effective alternative animal model has been developed using simian immunodeficiency virus (SIV) infection of macaques. This model has been extremely useful for investigating the pathogenesis of immunodeficiency virus infection and for the testing of vaccines and therapies (Gardner et al. [1994] in *The Retroviruses*, vol. 3 [Levy, ed.] Plenum Press, New York).

Experimental infection of nonhuman primates with human immunodeficiency virus type 1 (HIV-1) has been described for gibbons (Lusso et al. [1988] J. Immunol. **141**:2467-2473), pig-tailed macaques (Barre-Sinoussi et al., *supra*), and chimpanzees (Alter et al. [1984] Science **226**:549-552; Fultz et al. [1986] J. Virol. **58**:116-124). However, the development of disease (AIDS) has not been documented in these animals or in any animal species infected with HIV-1. For many years, chimpanzees have been a major focus in the development of animal models for HIV-1 infection and therapy. The ability to consistently infect chimpanzees with several HIV-1 subtypes and to reisolate the virus over extended periods (Fultz et al., *supra*; Johnson et al. [1993] AIDS Res. Hum. Retroviruses **9**:375-378) has made chimpanzees useful for testing vaccine candidates (Berman et al. [1990] Nature **345**:622-625; Fultz et al. [1992] Science **256**:1687-1690; Girard et al. [1995] J. Virol. **69**:6239-6248). However, the lack of disease in HIV-1-infected chimpanzees has raised concern over the relevance of these vaccine studies.

The present invention provides HIV-1 isolates that are infectious to nonhuman primates such as chimpanzees and induce AIDS in inoculated nonhuman primates including, but not limited to, chimpanzees. Further, this invention provides for the first time a proper animal model for HIV-1 infection and for the development of AIDS.

SUMMARY OF THE INVENTION

A first aspect of the present invention are the isolated human immunodeficiency virus type 1 (HIV-1) isolate having the identifying characteristics of HIV-1 isolate JC ($HIV\text{-}1_{JC}$) and assigned AIDS Reagent Program Catalog Number 3523 and the isolate having the identifying characteristics of HIV-1 isolate NC ($HIV\text{-}1_{NC}$). These viruses are useful for the preparation of recombinant, attenuated and subunit vaccines, as well as for the preparation of challenge stocks. Each is also useful in screening for the presence of HIV in biological samples.

A second aspect of the invention is a biological sample, e.g., a biological fluid or tissue, containing an HIV-1 having the identifying characteristics of $HIV\text{-}1_{JC}$ or $HIV\text{-}1_{NC}$. In a particular embodiment of the invention, primate blood specimens containing the $HIV\text{-}1_{JC}$ or the $HIV\text{-}1_{NC}$ of the present invention were used to cause HIV infection and to induce AIDS in a nonhuman primate, as specifically exemplified by the chimpanzee.

A third aspect of the invention are a biologically pure culture of host cells containing an HIV-1 having the identifying characteristics of $HIV\text{-}1_{JC}$ and a biologically pure culture of host cells containing an HIV-1 having the identifying characteristics of $HIV\text{-}1_{NC}$.

A further aspect of the invention are isolated DNA molecules which produce infectious HIV-1 having the identifying characteristics of $HIV\text{-}1_{JC}$ or which encode an antigenic fragment thereof or having the identifying characteristics of $HIV\text{-}1_{NC}$ or encode an antigenic fragment thereof. The substantially full length sequence for infectious molecular clone of $HIV\text{-}1_{JC}$ is given in SEQ ID NO:11, and the corresponding sequence for the infectious molecular clone of $HIV\text{-}1_{NC}$ is given in SEQ ID NO:12.

An additional aspect of the invention provides isolated DNA encoding an HIV-1 envelope protein having the amino acid sequence of SEQ ID NO:2. In a particular embodiment, the DNA sequence encoding an HIV-1 envelope protein contains the nucleotide sequence of SEQ ID NO:1.

A further aspect of the invention are a composition comprising an antigenic preparation derived from the HIV-1_{JC} of the invention and a composition comprising an antigenic preparation derived from the HIV-1_{NC} of the present invention.

An additional aspect of the invention are pharmaceutical compositions comprising an immunogenic amount of an antigenic preparation derived from the HIV-1_{JC} of the invention in a pharmaceutically acceptable carrier or HIV-1_{NC} in a pharmaceutically acceptable carrier.

A further aspect of the present invention is a kit for detecting the presence of HIV-1 antibodies, comprising an antigenic preparation derived from the foregoing HIV-1_{JC} or from HIV-1_{NC} as described herein.

Another aspect of the invention is a method of inducing in a subject antibodies to the HIV-1_{JC} or to HIV-1_{NC} of the invention, comprising the step of administering to the subject an immunogenic amount of an antigenic preparation derived from the HIV-1_{JC} or from the HIV-1_{NC}. The induced antibodies can be harvested and find use, for example, as hybridization probes in methods for HIV scanning.

A further aspect of the invention is a method of immunizing a subject against infection by HIV, comprising the step of administering to the subject an immunogenic amount of an antigenic preparation derived from the HIV-1_{JC} or derived from HIV-1_{NC} of the present invention.

It is another aspect of the invention to provide a method for inducing acquired immune deficiency syndrome (AIDS) in a nonhuman primate, comprising the step of administering to said primate an effective amount of an antigenic preparation derived from the HIV-1_{JC} and/or HIV-1_{NC} of the invention such that said primate develops AIDS. In a specific embodiment of the invention, HIV-1_{JC} induced AIDS in chimpanzees, e.g., in a chimpanzee referred to as C499. In a different embodiment of the invention, HIV-1_{JC} contained in a biological sample from a chimpanzee having AIDS (C499) induces AIDS in a second chimpanzee and in a different primate species.

It is another aspect of the invention to provide a method for inducing acquired immune deficiency syndrome (AIDS) in a nonhuman primate, comprising the step of administering to said primate an effective amount of an antigenic preparation derived from the HIV-1_{NC} of the invention such that said primate develops AIDS. In a specific embodiment of the invention, HIV-1_{NC} induced AIDS in chimpanzees, e.g., in a chimpanzee referred to as C534. In a different embodiment of the invention, HIV-1_{NC} contained in a biological sample from a chimpanzee having AIDS (C455) induces AIDS in a second chimpanzee and in a different primate species.

A further aspect of the invention provides a biological fluid or tissue sample, obtained from a chimpanzee having HIV-1_{JC}-induced AIDS, containing an antigenic HIV fragment for inducing AIDS in a nonhuman primate.

A further aspect of the invention provides a biological fluid or tissue sample, obtained from a chimpanzee having HIV-1_{NC}-induced AIDS, containing an antigenic HIV fragment for inducing AIDS in a nonhuman primate.

It is yet another aspect of the invention to provide a method for inducing AIDS in a nonhuman primate, comprising the step of administering to said primate an effective amount of a biological fluid or tissue sample, obtained from a primate having HIV-1_{JC}-induced or HIV-1_{NC}-induced AIDS, containing an antigenic HIV fragment for inducing AIDS in a nonhuman primate.

The present invention further provides various vaccine formulations containing active immunogenic agents derived from the foregoing HIV-1_{JC}, DNA encoding the HIV-1_{JC}, and DNA encoding antigenic fragments of the HIV-1_{JC}, from the foregoing HIV-1_{NC}, DNA encoding the HIV-1_{NC}, and DNA encoding antigenic fragments of the HIV-1_{NC}. An antigenic fragment contains one or more epitopes which bind antibodies directed to the HIV-1_{JC} and/or HIV-1_{NC} of the invention.

BRIEF DESCRIPTION OF THE FIGURES

Fig. 1 shows the CD4+ -cell decline and plasma virus loads in chimpanzee C455 following transfusion with blood from C499. Immediately before and at various times after transfusion, blood was collected from C455 for *in vitro* analyses of CD4+ -cell levels and plasma virus loads. Absolute peripheral CD4+ cells in C455 showed a dramatic decrease by 2 weeks post-transfusion. This rapid decline continued and, by 14 weeks after transfusion, the number of CD4+ cells decreased to 10/μl. These low cell numbers have been maintained to date (42 weeks post-transfusion). Plasma HIV-1 RNA loads in C455 showed high levels of virus present by two weeks after transfusion. Results obtained at the 4-week point suggest very high levels of virus replication in C455. However, HIV-1 levels appeared to be somewhat controlled by 5 to 9 weeks post-transfusion. The cutoff level of 10^4 equivalents/ml (---) is the lower limit of the assay.

Fig. 2 illustrates CD4+ -cell decline in chimpanzee C534 following transfusion with blood from C455. Immediately before and at various times after transfusion, blood was collected from C534 for *in vitro* analyses of CD4+ -cell levels. Absolute peripheral CD4+ cells in C534 showed a dramatic decrease by 20 days post-transfusion.

Figure 3 illustrates the strategy used in PCR amplification of subgenomic fragments from HIV-1 and the location and orientation of primers on the viral genome are shown. All the primers were designed from the HIV-1_{LAI} nucleotide sequence and their coordinates are described below. PCR amplified fragments were cloned in TA vectors and the corresponding name designations for the recombinant plasmids are shown in bracket. The 5'-LTR-containing *Apa I* fragment amplified from HIV-1_{JC} PBMC genomic DNA was subcloned in pJC to generate plasmid pHIV-1_{JC16} while the *ApaI-NcoI* fragment containing the 5'-LTR region amplified from HIV-1_{NC} PBMC genomic DNA was subcloned in pHIV-1_{JC16} to generate pHIV-1_{NCJC} chimeric plasmid. Plasmid pHIV-1_{NC7} was constructed by subcloning the *env*-containing *NcoI-XhoI* fragment amplified from HIV-1_{NC} genomic DNA to pHIV-1_{NCJC} chimeric plasmid. All recombinant plasmids (pHIV-1_{JC16}, pHIV-1_{NCJC}, and pHIV-1_{NC7}) lacked 55 bp at the 5' end of the genome (U3 region) and all of U5 region in the 3' LTR region.

Figures 4A-4F show replication of cloned and uncloned HIV-1 isolates in con-A stimulated and unstimulated chimpanzee PBMC (cPBMC). Stimulated cPBMC infected with (Fig. 4A) HIV-a LAV-1b and SF2 parental strains and the highly cytopathic DH12 isolate, (Fig. 4B) JC (uncloned) and JC16 (cloned) isolates of HIV-1, and (Fig. 4C) NC (uncloned) and NC7 (cloned) isolates of HIV. Unstimulated cPBMC infected with: (Fig. 4D) LAV-1b, SF2, and DH12; (Fig. 4E) JC, JC16; and (Fig. 4F) NC, NC7. Chimpanzee PBMC (1.1×10^7) in T-25 cm^2 were infected with 20 ng of either HIV-1_{NC} (uncloned), HIV-1_{SF2}, HIV-$1_{LAV\text{-}1b}$, or HIV-1_{DH12} virus and incubated for a total of 17 days at 37°C. Supernatant aliquots were made on 3, 7, 10, 14, and 17 days post infection. Reverse transcriptase (RT) assays were performed as outlined hereinbelow.

Figure 5 shows replication of HIV-1 isolates in chimpanzee monocyte-derived macrophages (MDM). Purified PBMC (6×10^6/well) were used to obtain MDM. Ten ng of each of virus HIV-1_{JC16} (molecular clone), HIV-1_{JC} (uncloned), HIV-1_{NC7} (molecular clone), HIV-1_{NC} (uncloned), HIV-1_{SF2}, HIV-$1_{LAV\text{-}1b}$, and HIV-1_{DH12}) was used for infection and on days 7 and 14 post infection supernatants were harvested. The mount of virus in the supernatants was determined using the p24 HIV-1 antigen capture ELISA (Coulter). Supernatants from control uninfected cultures are represented.

DETAILED DESCRIPTION OF THE INVENTION

The following definitions are given in order to provide clarity as to the intent or scope of their usage in the specification and claims.

The term HIV-1 isolate JC or HIV-1_{JC} or HIV_{JC} or HIV-1 having the identifying characteristics of HIV-1_{JC} as used herein refers to the particular HIV-1 isolated from a chimpanzee (C499) that developed AIDS 10 years after infection with HIV-1 (Novembre et al. [1997] J. Virol. 71:4086-4091). The DNA sequence of the cloned HIV-1_{JC} is given in SEQ ID NO:11.

The term HIV-1 isolate NC or HIV-1_{NC} or HIV_{NC} or HIV-1 having the identifying characteristics of HIV-1_{NC} as used herein refers to the particular HIV-1 isolated from a chimpanzee (C455) that developed AIDS after infection with HIV-1_{JC} (Novembre et al. [1997] J. Virol. 71:4086-4091). The DNA sequence of the cloned HIV-1_{NC} is given in SEQ ID NO:12.

The term antigenic preparation of HIV-1_{JC} or antigenic fragment of HIV-1_{JC} as used herein refers to the whole viral particle of the HIV-1_{JC} or to a fragment thereof, wherein such fragment encodes at least one epitope or antigenic determinant. The term antigenic preparation of HIV-1_{NC} or antigenic fragment of HIV-1_{NC} as used herein refers to the whole viral particle of the HIV-1_{JC} or to a fragment thereof, wherein such fragment encodes at least one epitope or antigenic determinant.

The term animal or subject as used herein refers to a mammal and, more frequently, to a primate.

The term effective amount as used herein refers to the quantity of active ingredient necessary to effect in an animal a change in a specific biochemical or immunological parameter. For example, in a particular embodiment of this disclosure, an effective amount refers to the amount of HIV-1 administered to a subject such that viral infection and AIDS developed.

The term antigenic amount as used herein refers to the quantity of active antigen necessary to effect an interaction with corresponding antibodies.

The term immunogenic amount as used herein refers to the quantity of active antigen necessary to stimulate the immune system in response to a specific antigen.

The term labeled is used herein to refer to the conjugating or covalent bonding of any suitable detectable group, including enzymes (e.g., horseradish peroxidase, β-glucuronidase, alkaline phosphatase, and β-D-galactosidase), fluorescent labels (e.g., fluorescein, luciferase),

and radiolabels (e.g., ^{14}C, ^{131}I, ^{3}H, ^{32}P, and ^{35}S) to the compound being labeled. Techniques for labeling various compounds, including proteins, peptides and antibodies are well known. See, e.g., Morrison, *Methods in Enzymology* 32B, 103 (1974); Syvanen et al., *J. Biol. Chem.* 284, 3762 [1973]; Bolton and Hunter, *Biochem. J.* 133, 529 [1973].

The viruses of the present invention resulted from the development of AIDS in a first chimpanzee infected with HIV-1 for over 10 years and by the rapid development of immunosuppression in a second chimpanzee transfused with blood from the first chimpanzee. To investigate the ability of HIV-1 to induce AIDS in non-human primates, a cohort of 12 chimpanzees was inoculated with several strains of HIV-1 at the Yerkes Center in the mid-1980s. A member of this cohort, C499, was described as part of a previously reported superinfection study (Fultz et al. [1987] J. Virol. **61**:4026-4029) and was inoculated on three separate occasions with three different HIV-1 isolates: HIV-1_{SF2} in 1985, HIV-1_{LAI} in 1986, and HIV-1_{NDK} in 1987. The first inoculation resulted in infection, as determined by positive virus isolation and persistent HIV-1-specific antibody response. Clinically, the animal remained healthy except for the development of thrombocytopenia and lymphopenia in 1988, which resolved without treatment (Fultz et al. [1991] J. Infect. Dis. **163**:441-447). (All HIV-infected chimpanzees are maintained in Biosafety level 3 isolation facilities.)

In 1993, after the resumption of yearly monitoring (monitoring of all HIV-1-infected chimpanzees at the Yerkes Center was suspended from March 1990 to May 1993), a decrease in the levels of platelets and $CD4^+$ cells in C499 was observed (24,000 and 390/µl, respectively) (Table 1). Thrombocytopenia and $CD4^+$-cell lymphopenia were persistent in this animal from this point onward. In addition, C499 displayed other significant clinical signs of disease. Beginning in March 1995, C499 developed chronic, intermittent diarrhea for which no enteric pathogens were identified and which was not resolved with antibiotic treatments. In September 1995, this animal developed acute fulminant diarrhea which was associated with large numbers of *Blastocystis hominis* and *Balantidium coli*. At this time, $CD4^+$ cells decreased to extremely low levels (minimum of 10/µl:2% of total T cells) (Table 1), indicative of severe immunosuppression. Similarly, declines were observed in the levels of total lymphocytes and $CD8^+$ cells (Table 1). Treatment with fluid replacement and

antimicrobial and antiprotozoal therapy (doxycycline, ceftriaxone, enrofloxacin, gentamicin, and albendazole) resulted in the resolution of acute diarrhea within five days.

Because this was the first chimpanzee to develop AIDS, treatment with antiretroviral therapies was not administered in order to more fully characterize virological, immunological, and pathological parameters in this animal. As the acute diarrhea was resolved with treatment, $CD4^+$-cell levels rose to a maximum of 180/μl (4% of total T cells) but declined again. Subsequently, chronic, intermittent diarrhea resumed and continued unresolved. During this period, C499 exhibited no lymphadenopathy or wasting. However, beginning in the latter part of 1995 and extending into 1996, C499 developed progressive nonregenerative anemia (hematocrit levels of 37.5% in December 1995 and 27.2% in January 1996, with respective hemoglobin values of 12.1 and 8.4 g/dl; reticulocyte counts were 0.0% since November 1995). Due to progressive hematologic abnormalities, chronic diarrhea, and continued immunosuppression, the animal was euthanized in February 1996.

Concomitant with the decrease in $CD4^+$ cells was an increase in HIV-1 loads in plasma (Table 1). The increase in the level of the virus was detected in plasma samples dating from May 1993, when the $CD4^+$-cell decline was first noted, but not before the suspension of monitoring in 1990. These levels are significantly higher than those for five other chimpanzees at the Yerkes Center which received cell-free or cell-associated HIV-1_{LAI} or HIV-1_{SF2} inoculations (all have undetectable plasma viral RNA levels and are not immunosuppressed). These results suggest that pathogenic effects began to occur sometime between 1990, when $CD4^+$-cell counts were at normal levels and viral loads were undetectable, and 1993, when alterations in the number of $CD4^+$ cells and significant virus loads were present.

The ability to isolate virus from C499 varied since C499's first exposure to HIV-1. In general, early after inoculation, HIV was easily isolated from the peripheral blood mononuclear cells (PBMC) of this animal; however, after several months, virus could no longer be isolated. From August 1988 until the suspension of monitoring in 1990, the virus was consistently isolated from C499 on a monthly basis. Subsequently, after the resumption

of monitoring in mid-1993, HIV-1 continued to be easily isolated from this animal. Quantitative titration of PBMC viral load following the development of acute diarrhea in C499 in September 1995 until the time of euthanasia revealed that 10^4 to 10^5 PBMC was consistently required for virus isolation. The immune response of C499 to HIV-1 infection was very strong up to the time of euthanasia. HIV-1 antibody endpoint titers (HIV-1 whole-virus enzyme-linked immunosorbent assay, Genetic Systems, Redmond, WA) ranged from 51,200 to 204,800 since 1993. Because of the deteriorating condition of this animal and because of the severe decline in $CD4^+$ cells, it was hypothesized that the HIV-1 present in this animal had evolved to become more cytopathic for chimpanzee $CD4^+$ cells. Cocultivation of PBMC derived from C499 (obtained at the time of acute diarrhea) with uninfected chimpanzee PBMC (cPBMC) resulted in the isolation of a virus (HIV-1_{JC}) which induced syncytium formation in chimpanzee cells (Fig. 1A-D). This characteristic has been previously described for only three HIV-1 isolates (Ghosh et al. [1993] Virology **194**:858-864; Schuitemaker et al. [1993] J. Infect. Dis. **168**:1140-1147; Shibata et al. [1995] J. Virol. **69**:4453-4462), none of which were used for inoculation of C499.

The virus strain derived from C499 (HIV-1_{JC}) and other HIV-1 isolates (HIV-1_{LAI} and HIV-1_{SF2}) were tested for the ability to induce syncytium formation in cPBMC. Virus stocks were prepared in cPBMC (HIV-1_{JC}) or in human PBMC (HIV-1_{LAI} and HIV-1_{SF2}). Cells were incubated with virus overnight and were then washed. Cultures were examined daily for evidence of cytopathic effects. cPBMC four days after infection with an HIV-1 isolate (HIV-1_{JC}) from C499 showed beginnings of syncytium formation and separated syncytia. cPBMC four days after infection with HIV-1_{LAI} lacked of syncytium formation, and normal cell clusters were present. cPBMC four days after infection with HIV-1_{SF2} also showed no syncytium formation. The only virus to induce significant cytopathic effects in cPBMC was HIV-1_{JC}. All cultures were examined for 14 days following infection. All cultures, regardless of the virus used, became positive for virus replication by seven days postinfection. Thus, it is concluded that genetic changes which confer the ability to induce syncytium formation occurred in the virus present in C499.

To confirm that the virus present in C499 was different from the viruses used for inoculation, DNA prepared from HIV-1_{JC}-infected cPBMC was used as a template in typical PCR assays with primers (forward, no. 384: 5'CCCTTCGAAGAGGATATAATCAGTTTATGGGATCAAAGC3' [SEQ ID NO:9]; reverse, no. 383: 5'CCCTTCGAACTCTTCTTCTGCTAGACTGCCATT3' [SEQ ID NO:10]) designed to amplify a 507-bp fragment of the *env* gene containing the V1 and V2 regions. Genetic analysis of 16 HIV-1_{JC} V1-V2 clones obtained by ligation of the amplification products with the vector pGEM7ZF (Promega, Madison, WI) showed amino acid homologies of 80 to 84% with HIV-1_{LAI}, 73 to 80% with HIV-1_{SF2}, and 63- to 68% with HIV-1_{NDK}. Thus, there appears to be considerable divergence between the virus present in C499 at the time of acute disease and the viruses used to inoculate this animal. This divergence is further illustrated in Fig. Tables 3A-3C, which show amino acid alignments of five HIV-1_{JC} clones, HIV-1_{NDK}, HIV-1_{LAI}, and HIV-1_{SF2}. Comparative analyses between the 16 HIV-1_{JC} clones showed that amino acid homologies ranged from 81 to 96%, with no clones being identical. These results suggest that the virus population in C499 consisted of a large quasispecies. Furthermore, the data, when combined with the *in vitro* analyses described above, indicate that the virus adapted after years of replication and mutation, becoming more pathogenic for the chimpanzee. While no evidence of recombination is evident from analyses performed in this small area, the possibility of recombination cannot be ruled out for other portions of the HIV-1_{JC} genome.

Tables 2A-2C illustrate amino acid alignment of V1-V2 clones obtained from HIV-1_{JC}-infected cPBMC and prototypes HIV-1_{LAI} and HIV-1_{SF2}, and HIV-1_{NDK}. *env* clones of HIV-1_{JC} encompassing the V1-V2 region were sequenced by the dideoxy chain termination method (Sequenase; Amersham Life Science, Arlington Heights, IL). With the Intelligenetics Suite of programs (Intelligenetics, Beaverton, OR), sequences of HIV-1_{JC} *env* fragments were used to derive corresponding amino acid sequences. Deduced amino acid sequences from five of these clones ($HIV_{JC}10$, $HIV_{JC}17$, $HIV_{JC}45$, $HIV_{JC}48$, and $HIV_{JC}55$) were then aligned with the corresponding region in HIV-1_{LAI}, HIV-1_{SF2}, and HIV-1_{NDK} isolates. Amino acid sequences of HIV-1_{LAI} ,HIV-1_{SF2} and HIV-1_{NDK} were obtained from the Human Retroviruses and AIDS Database.

To investigate additional pathogenic effects of HIV-1 infection on C499, tissue samples obtained by biopsy during the acute diarrheal stage (September 1995) and at necropsy were subjected to histopathological analyses. First, examination of a peripheral lymph node (obtained from C499 in September 1995) revealed marked lymphoid depletion within the cortical area, with a few follicles remaining, as compared with a lymph node from an age-matched, uninfected chimpanzee, which lacked follicular development and had a very cellular cortex. To detect virus expression in tissues, *in situ* hybridization experiments with digoxigenin-labeled riboprobes encompassing the entire HIV-1 genome (derived from the HIV-1 BH10 molecular clone (Hahn et al. [1984] Nature **312**:166-169) were used to probe formalin-fixed lymph node sections. Detection of bound probes was performed with a sheep anti-digoxigenin-alkaline phosphatase-labeled Fab monoclonal antibody and with Nitro Blue Tetrazolium-5-bromo-4-chloro-3-indolylphosphate toluidinium as the substrate chromogen, according to previously described methods (Hirsch et al. [1995] J. Virol. **69**:955-967). *In situ* hybridization studies of C499's lymph node demonstrated the presence of HIV-1 RNA in follicular trapping patterns, with few positively staining cells. These findings are similar to changes observed in HIV-1-infected persons developing AIDS (Pantaleo et al. [1995] N. Engl. J. Med. **332**:209-216). Control samples, which included sense riboprobes and lymph nodes from uninfected chimpanzees, did not show any staining. Additional lymph nodes obtained from C499 at necropsy showed a similar depleted pattern. However, a few lymph nodes were not as depleted but did contain multinucleated giant cells which stained positive for the HIV-1 p24 antigen. These giant cells are often found in the lymph nodes of simian immunodeficiency virus (SIV)-infected macaques and are occasionally found in the lymph nodes in HIV-1-infected persons.

Finally, histopathologic analysis of intestinal tissue from C499 revealed pathologic changes in the ileum, with significant blunting of the villi and intense infiltration of mononuclear cells and plasma cells. Examination of the intestinal mucosa at higher magnification revealed extensive infection with *Cryptosporidium*, which was present throughout the small intestine, lining the apical surfaces of intestinal epithelial cells. This organism is an AIDS-defining opportunistic pathogen (Centers for Disease Control and Prevention. 1992. 1993 revised classification system for HIV infection and expanded

surveillance case definition for AIDS among adolescents and adults. Morbid. Mortal. Weekly Rep. 41:1-19) and probably accounted for the chronic intermittent diarrhea and intestinal pathology in C499.

Most tissues obtained from C499 following euthanasia appeared grossly normal, with no lymphadenopathy or splenomegaly. However, the spleen showed moderate congestion and lacked follicle development. Bone marrow specimens obtained at euthanasia showed some functional impairment in CFU granulocyte macrophage and CFU formation, although levels of $CD4^+$ cells were not altered (Villinger et al. [1997] J. Med. Primatol **26**:181-189. Virus was isolated from all lymphoid organs including inguinal, axillary, and mesenteric lymph nodes as well as from the spleen and thymus. Virus was also isolated from the kidney and liver but not from the brain or cerebrospinal fluid.

At the time of acute diarrhea in C499, a blood transfusion was performed to determine the effects of passage of virus from this animal to an uninfected chimpanzee. Forty milliliters of blood obtained by venipuncture from C499 was immediately transfused intravenously into C455. This chimpanzee, which was bred in captivity, was seropositive for Epstein-Barr virus and cytomegalovirus and was seronegative and PCR negative for HIV prior to the transfusion. Results of titration analysis of PBMC and plasma from C499 show that in the 40 ml of blood, 1×10^4 50% tissue culture infective doses ($TCID_{50}$) of virus was in PBMC and 2×10^4 $TCID_{50}$ of virus was in plasma. Thus, C455 received a total of 3×10^4 $TCID_{50}$ of virus. Analysis of peripheral $CD4^+$ cell levels in C455 revealed a precipitous decline beginning by 2 weeks after transfusion, when absolute numbers of $CD4^+$ cells decreased from 1,240 to 320 cells/μl (Fig. 1). This decline continued, reaching a minimum value of 10 cells/μl (1% of total T cells) by 14 weeks post-transfusion. Since this time, the total percentage of T cells that the $CD4^+$ population encompasses has remained constant at 1%, with only a slight rise in the absolute number of $CD4^+$ cells (20 cells/μl at the latest time point). Levels of peripheral $CD8^+$ cells showed an initial decline from 1,590 to 880 cells/μl in the first two weeks after transfusion. However, these levels quickly rebounded to 2,010 $CD8^+$ cells/μl by 8 weeks after transfusion. Since this time, the number of $CD8^+$ cells in circulation has been maintained between 550 and 4,320/μl, with the level in most recent sample being 840 cells/μl.

Because $CD4^+$-cell levels declined so rapidly, the level of virus present in C455 was investigated by quantitation of plasma HIV-1 RNA with the Chiron B-DNA assay (Fig. 1). Two weeks after transfusion, plasma viral loads were 2×10^7 RNA equivalents/ml and reached maximum levels by 4 weeks post-transfusion (6.2×10^7 RNA equivalents/ml). Following a decline (which corresponded with the development of anti-HIV-1 specific antibody), plasma HIV-1 levels have been maintained at ~1.1×10^5 RNA equivalents/ml. Virus has been easily isolated from C455 at all times post-transfusion. Quantitative coculture of PBMC from C455 has shown that early after transfusion, virus could be isolated from as few as 10^2 PBMC, while at more recent time points, up to 10^5 to 10^6 cells was required for virus isolation. *In vitro* analysis of virus isolated from this animal at several time points has shown that the ability to induce syncytium formation in cPBMC has been retained, further implicating this virus in the pathogenesis of $CD4^+$-cell decline. Antibody responses to HIV-1 have been moderate in C455. Enzyme-linked immunosorbent assay titers have been maintained between 1,600 and 6,400 since four weeks post-transfusion. Clinically, C455 has appeared normal, except for episodic incidences of a rash on the chest and in the scrotum area. In addition, the animal has experienced no weight loss, lymphadenopathy, or anemia.

This invention presents an HIV-1 infected chimpanzee that developed AIDS as defined by the Centers for Disease Control and Prevention classification system (CDCP 1992, *supra*). Progression of clinical disease (anemia, thrombocytopenia, and chronic diarrhea) in this animal was associated with several key findings, including the following: (i) the presence of a virus which is cytopathic for cPBMC *in vitro* and *in vivo* and is genetically distinct from those used for inoculation, (ii) an increase in viral load; (iii) $CD4^+$-cell depletion, (iv) lymph node depletion, and (v) the presence of *Cryptosporidium* organisms in the intestine. It appears that the critical change(s) associated with clinical progression may have developed during the period in which C499 was not monitored (mid-1990 through mid-1993). The precipitous $CD4^+$-cell decline concomitant with high viral loads displayed in C455, transfused with blood from C499, suggest that the HIV present in C499 has evolved to become more pathogenic for chimpanzees. The increased pathogenicity of a lentivirus after passage into a new host has been previously observed with the adaptation of SIV from sooty

mangabeys to pig-tailed macaques, resulting in the development of the acutely lethal strain SIVsmmPBj14 (Fultz et al. [1989] AIDS Res. Hum. Retroviruses 5:397-409).

Forty milliliters of blood obtained by venipuncture from C455 was immediately transfused intravenously into C534. This chimpanzee, which was bred in captivity, was seropositive for Epstein-Barr virus and cytomegalovirus and was seronegative and PCR negative for HIV prior to the transfusion. Results of titration analysis of PBMC and plasma from C455 show that in the 40 ml of blood, 1 x 10^4 50% tissue culture infective doses ($TCID_{50}$) of virus was in PBMC and 2 x 10^4 $TCID_{50}$ of virus was in plasma. Thus, C534 received a total of 3 x 10^4 $TCID_{50}$ of virus. Analysis of peripheral $CD4^+$ cell levels in C534 revealed a precipitous decline beginning by one week after transfusion, when absolute numbers of $CD4^+$ cells decreased from approximately 1,450 to 700 cells/μl (Fig. 2). The total percentage of T cells that the $CD4^+$ population encompasses has remained constant at approximately 1%, with only a slight rise in the absolute number of $CD4^+$ cells (20 cells/μl at the latest time point). Levels of peripheral $CD8^+$ cells showed an initial decline in the first two weeks after transfusion. However, these levels quickly rebounded after transfusion.

Table 3 shows the antibody response in chimpanzee C534 following transfusion with blood from C455. Immediately before and at various times after transfusion, blood was collected from C534 for *in vitro* analyses of antibody response. Plasma HIV-1 RNA loads in C534 showed high levels of virus present by two weeks after transfusion. Results obtained at the 3-week point indicate very high levels of virus replication in C455. The cutoff level of 10^4 equivalents/ml is the lower limit of the assay.

Because $CD4^+$-cell levels declined so rapidly, the level of virus present in C534 was investigated by quantitation of plasma HIV-1 antibody response (Table 4). One week after transfusion, the plasma antibody titer was still 0. At two weeks after transfusion, the antibody titer rose to 400 and remained at 1600 at three and four weeks post-transfusion. Virus has been easily isolated from C455 at all times after transfusion. Quantitative coculture of PBMC from C534 has shown that early after transfusion, virus could be isolated from as few as 10^2 PBMC, while at more recent time points, up to 10^5 to 10^6 cells was required for virus

isolation. *In vitro* analysis of virus isolated from this animal at several time points has shown that the ability to induce syncytium formation in cPBMC has been retained, further implicating this virus in the pathogenesis of $CD4^+$-cell decline. Antibody responses to HIV-1 have been moderate in C534. Enzyme-linked immunosorbent assay titers have been maintained between 1,600 and 6,400 since four weeks post-transfusion. Clinically, C534 has appeared normal, except for episodic incidences of a rash on the chest and in the scrotum area. In addition, the animal has experienced no weight loss, lymphadenopathy, or anemia (normal hematocrit levels).

This invention presents HIV-1 isolates NC and JC which infected chimpanzees such that the HIV-1 infected chimpanzees develop AIDS as defined by the Centers for Disease Control and Prevention classification system (CDCP 1992, *supra*). Progression of clinical disease (anemia, thrombocytopenia, and chronic diarrhea) in this animal is associated with several key findings, including the following: (i) the presence of a virus which is cytopathic for cPBMC *in vitro* and *in vivo* and is genetically distinct from that used for inoculation, (ii) an increase in viral load; (iii) $CD4^+$-cell depletion, (iv) lymph node depletion, and (v) the presence of *Cryptosporidium* organisms in the intestine. It is expected that the critical change(s) associated with clinical progression develop during the period spanning approximately 3 to 5 years after HIV-1 infection. The precipitous $CD4^+$-cell decline concomitant with high viral loads displayed in C534, transfused with blood from C455, indicates that the HIV present in C455 remains pathogenic for chimpanzees.

In the past, relevance of the HIV-1-infected chimpanzee as a model for vaccine evaluation was questioned due to lack of disease development. The lack of an animal model which supports pathogenic HIV-1 infection has been a continuing problem for vaccine development. Although the time of progression to disease (>10 years), the currently limited numbers of animals available for use, and the overall high costs associated with working with chimpanzees are deterrents to their widespread use in AIDS research, the potential usefulness of this model cannot be disregarded. The development of AIDS in C499, the fact that additional HIV-infected chimpanzees have depressed $CD4^+$-cell counts (<500/μl) and thrombocytopenia, and the rapid progression of the $CD4^+$-cell decline in C455 support the

role that this animal model provides in AIDS-related studies. The adaptation of HIV-1 from long-term chimpanzee infection to a pathogenic form provides a critical link for the adaptation of HIV-1 to growth in more readily available nonhuman primate species. The instant invention further contemplates the growth of HIV-1_{JC} isolated from C499 and/or HIV-1_{NC} from C455 in chimpanzees as well as in pig-tailed or rhesus macaques. In addition, the present invention provides for continued biological and genetic characterization of HIV-1_{JC} and HIV-1_{NC} with further key insights into the pathogenesis of HIV-1 infection in humans and chimpanzees, for example, for the development of drugs and vaccines for the treatment and prevention of AIDS.

As described above, chimpanzee C499 was initially infected with HIV-1_{SF2} in 1985 [Fultz et al. (1986) *J. Virol.* 58, 116-124] and later inoculated with HIV-$1_{LAV\text{-}1b}$ and HIV-1_{NDK} in 1986 and 1987 respectively. Superinfection with HIV-$1_{LAV\text{-}1b}$ (but not HIV-1_{NDK}) was demonstrated by restriction enzyme analysis of PBMC genomic DNA. At the time of disease development in C499, a virus isolate, termed HIV-1_{JC}, was obtained by co-culture of C499 PBMC with normal cPBMC. At that time, sequence analysis of the V1-V2 region of env suggested that HIV-1_{JC} was most closely related to HIV-1_{LAV}. Also at the time of disease development, blood from C499 was transfused into an uninfected chimpanzee (C455) which resulted in a dramatic decline of CD4+ cells by 2 weeks post transfusion. The depressed CD4+ cell count is still maintained to date in this animal. One month post transfusion 50 µl of plasma from C455 was used for in vitro infection of normal chimpanzee PBMC and the resultant virus was designated HIV-1_{NC}.

To perform a more thorough analysis of the genetic makeup of the HIV-1_{JC} and HIV-1_{NC} viruses, we constructed substantially full-length, infectious molecular clones as described in the Examples and in Figure 3. Both clones lacked 55 nucleotides at the 5' end (5' LTR, U3 region) and all of the U5 region in the 3' LTR. Several clones (representing both HIV-1_{JC} and HIV-1_{NC}) which appeared to be the correct size were tested for biological activity by transfection of CEMx174 cells. Supernatants from transfected cells were used in RT assays to monitor virus production. Two clones, one from each group (HIV-1_{JC} [JC16] and HIV-1_{NC} [NC7]), were positive by RT and also showed massive syncytia formation (2 to 3 days post

transfection), similar to that observed with uncloned virus. To prepare stock viruses for use in in vitro assays, 293 cells were transfected with molecularly cloned DNAs followed by amplification with cPBMC as outlined hereinbelow.

The complete nucleotide sequences of JC16 and NC7 were determined as described in Example 7, and the sequences are given in SEQ ID NO:11 and SEQ ID NO:12, respectively. The genomes were determined to be 9193 nt (JC16) and 9196 nt (NC7) in length and contained open reading (ORFs) for all HIV-1-specific structural, regulatory and accessory genes. Alignment of JC16 and NC7 DNA sequences revealed that the two genomes were very similar, but contained a number of nucleotide changes spread throughout the genome.

compared with the parental inoculating viruses, and with gaps introduced to optimize alignment and treated as mismatches, the LTR sequences of JC16 and NC7 had percent nucleotide identities of 92.1% (LAV), 93.9% SF2), and 89.1% (NDK)–lower than that The most divergent region between NC7 and JC16 was the V5 region of the env gene. JC16 contained a 6 bp deletion in the gag gene relative to NC7 while NC7 had 3 bp deletion in the env gene region relative to JC16. In the LTR region there was a 98.7% nucleotide identity between JC16 and NC7 with all the changes being localized to the U5 region (Table 4).

Comparative analyses were performed between JC16, NC7, and the parental inoculating viruses, SF2, LAV, and NDK (The sequence of LAV-1b has not yet been determined). When observed upon direct comparison of JC16 and NC7 (Table 3). Most of the host/virus transcription binding factor sequences (sites for AP-1, NF-AT, NF-kB and Sp-1) and the TAR CORE and the Lys-tRNA sites were conserved (or had single point mutations) between the parental (SF2 and LAV strains) and progeny viruses. However, there were 3 point mutations unique to JC16 and NC7 at the NRF/NRE binding site.

The deduced amino acid sequences for all proteins of HIV-1_{JC} and HIV-1_{NC} were generated using the Intelligenetics suite of programs. Using the Lasergene program (DNASTAR Inc., Madison, WI), multiple alignments of all proteins were constructed to examine similarities between HIV-1_{NC}, HIV-1_{JC} and the parental viruses (Table 4). Based upon percent homology calculations, Gag, Pol, Vif, Tat, Rev, Env, and Vpu of JC16 and NC7

were most closely related to LAV, with vpr and Nef being most closely related to SF2. In no case was it apparent that the NDK isolate was the origin of a protein sequence. While most changes involved amino acid point mutations, several proteins of JC16 and NC7 contained amino acid insertions or deletions relative to the parental inoculating strains. A closer analysis of amino acid alignments revealed that the percent homologies could be misleading with regards to the origin of the protein. For example, in Tat, JC16 and NC7 were more homologous to LAV than to SF2 (based upon percent homologies). However, the JC16 and NC7 Tat proteins contained 15 amino acid deletions with respect to LAV--similar to that present in the SF2 isolate. Similar findings were observed in Gag and Pol.

Table 5 shows amino acid alignments of Gag, Nef, and Env proteins from JC16, NC7 and the inoculating viruses, LAV, SF2, and NDK. The deduced amino acid sequences for the Gag (a), Nef (b), and Env proteins of HIV-1 were determined using the Intelligenetics Suite of programs (Intelligenetics, Beaverton, OR) and the CLUSTAL method (DNASTAR Inc., Madison, WI) was used for the alignment of proteins. The virus isolates are shown by the designations at the beginning of every line as LAI (HIV-1_{LAV}), JC16 (HIV-1_{JC16}), NC7 (HIV-1_{NC7}), and SF2 (HIV-1_{SF2}). Dashes (-) denote amino acid deletion while dots (.) denote amino acid identity. The numbers after the amino acid sequence on the right show the position of the right-most amino acid in the line. The functional domains on Gag (a) and Nef (b) are indicated. MBD (a) refers to membrane-binding domain while MTD (b) refers to membrane-targeting domain. The hypervariable regions in the Env glycoprotein (c) are shown as V1 to V5 and the CD4 binding domain is shown above the sequence. Gp120 (SU) is the outer surface membrane Env glycoprotein. The NDK Env, LAI Env, and SFS Env protein sequences are given in SEQ ID Nos: 3, 4 and 5, respectively. The JC16 and NC7 Env protein sequences are given in SEQ ID Nos: 2 and 23, respectively. There were 7 (JC16 and LAI sequences) and 5 (NC7 sequence) amino acid deletions in the matrix protein (p17) relative to SF2 isolate sequence. At the C-terminal end of Gag polyprotein, the progeny viruses and SF2 virus had 12 amino acid deletions in p6 protein relative to LAV virus sequence (Table 5). However, the capsid (p24) and nucleocapsid (p7) proteins were generally well conserved including the cysteine residues within the zinc-finger domains. Point mutations unique to JC16 and NC7 were present in p17, p24, p7 and p6 peptides. The Lck binding domain

(proline-rich region) within Nef was well conserved with only one point mutation in LAI (Table 5). However, there were 4 amino acid insertions in the SF2 sequence relative to the other viruses at the N-terminal portion of Nef. Sequence analysis of eight other non-infectious clones (4 from JC and 4 from NC) confirmed observations made for the Gag and Nef deletions suggesting that these characteristics are a general property of the viruses obtained from C499 and C455.

The vast majority of point mutations, deletions, and insertions in these clones, relative to the parental viruses, were found in the env glycoprotein region (Tables 4 and 5). Alignment of the Env proteins revealed that all the cysteine residues resident in the protein were conserved between the viruses. There were a total of 30 predicted N-linked glycosylation sequences (Asn-X-Thr or Asn-X-Ser) for the parental strains compared to 27 (NC7) and 29 (JC16) for the progeny viruses. Most of the glycosylation sites reside in the SU portion of Env with HIV-1_{SF2} containing the highest number and HIV-1 isolate NC the least (25 for SF2, 24 for LAI and JC, and 23 for NC). Glycosylation of glycoproteins has been shown to influence the immune response toward virus infection. As expected, the gp120 (SU) glycoprotein contained the highest number of mutations. The V1-V2, V3, V4, and V5 hypervariable regions contained 10, 8, 8, and 3 point mutations respectively specific to JC16 and NC7. The V1-V2 region of JC16 and NC7 also contained insertions relative to the other viruses, resulting in amino acid lengths of 72 (progeny viruses), 69 (LAV), 70 (SF2), and 61 (NDK). There were multiple amino acid deletions in the V4 regions and single amino acid insertions in V5 regions of JC16 and NC7 relative to the parental strains. Both the CD4-binding domain and the proteolytic cleavage site (REKR) at the SU/TM junction were perfectly conserved.

The V3 region was the most interesting of the hypervariable regions. While the parental strains contained only 9 basic amino acids (Arg, Lys, and His positively charged residues) JC16 and NC7 HIV-1 isolates had 12 basic and 2 negatively charged residues (Asp and Glu). This gave the progeny viruses a net positive charge of 10 in the entire V3 region and an overall positive charge of +1 (LAV) or +2 (SF2) compared to the parental

strains. Eight of the 10 resultant positive charges for JC and NC isolates are located between residues 10 and 27 of V3 (Fig. 3) compared with 5 of 9 (LAV) and 4 of 9 (SF2) in the same region. At least within this region, JC and NC isolates seem to have a relatively high net positive charge of 4 (relative to SF2) and 3 (relative to LAV). Other researchers have shown that changes in basic amino acids in the middle portion of V3 loop (residues 10 to 27 in Table 5) can alter the syncytium-inducing properties and phenotype of the virus [Bhattacharyya et al. (1996) *AIDS Res. Hum. Retroviruses* 12, 83-90; De Wolf et al. (1994) *AIDS Res. Hum. Retroviruses* 10, 1387-1400; Okada et al. (1994) *AIDS Res. Hum. Retroviruses* 10, 803-811].

To study the biological activities of the cloned and uncloned viruses derived from C499 and C455, we conducted in vitro replication studies. Figures 4A, 4B and 4C shows the results of replication studies in mitogen-stimulated cPBMC. Two of the three viruses used for inoculation of C499, SF2 and LAV-1b, were able to replicate in stimulated cPBMC, albeit with different kinetics (Fig. 4A). The SF2 isolate grew very slowly and to low titers in cPBMC. In contrast, the LAV-1b isolate grew very well and with rapid kinetics, with a high titer of virus already present by day 7 post infection. The SF2 isolate was unable to induce detectable syncytium formation in cPBMC. The LAV-1b isolate, under these conditions, induced very few syncytia, in contrast to previously reported results [Watanabe et al. (1991) *J. Virol.* 65, 3344-3348]. Included in these analyses was the DH12 isolate of HIV-1 [Shibata et al. (1995) *J. Virol.* 69, 4453-4462]. This primary isolate from a human has been shown to be highly cytopathic for cPBMC. While this virus quickly established infection in the stimulated cell population, it did not grow to high titers. The numerous syncytia formed infection with DH12 could account for the lack of growth observed. The uncloned and cloned viruses of JC (Fig. 4B) and NC (Fig. 4C) replicated to levels comparable to those of LAV-1b for the same period of time. The rates of replication for cloned and uncloned viruses were indistinguishable.

The ability of HIV-1 virus isolates to replicate in unstimulated cPBMC was similarly evaluated. Results of this assay (Fig. 4D, 4E and 4F) showed that only the NC (cloned and uncloned) and the LAV-1b isolates of HIV-1 were capable of significant replication in unstimulated cPBMC. Replication rates for these viruses were less than 10% those observed

in stimulated cPBMC. Additionally, the kinetics of virus production in unstimulated cells was much slower than that observed in stimulated cPBMC. Interestingly, JC (cloned and uncloned) isolates of HIV-1 failed to replicate in cPBMC, reflecting an inherent biological difference between the JC and NC viruses. While the DH12 isolate was able to replicate in unstimulated cells, the levels of virus achieved were much less than that of the other viruses. The SF2 isolate was unable to replicate in unstimulated PBMC. Virus recovered from the molecular clones displayed the intrinsic replicative properties exhibited by the viruses they were derived from.

Reports of HIV-1 isolates able to replicate in chimpanzee macrophages has been controversial. To investigate the ability of these viruses and their respective clones to replicate in macrophages, we conducted in vitro assays using purified chimpanzee monocyte-derived macrophages (MDM). Figure 5 shows virus production in macrophages at 7 and 14 days post infection. Both the HIV-1 JC and NC virus isolates (cloned and uncloned) replicated in MDM as determined by the levels of p24 antigen produced. However, the amount of virus produced by the HIV-1_{JC16} molecular clone at 14 days post infection was twice that produced by the HIV-1_{JC}. The titers for cloned and uncloned HIV-1_{NC} viruses were comparable for the same period of time. Among the viruses that were tropic for MDM, HIV-1_{JC} (uncloned) produced the least amount of virus. HIV-$1_{LAV\text{-}1b}$ and HIV-1_{DH12} also infected MDM and produced virus which is consistent with previous observations made by Gendelman et al. [Gendelman et al. (1991) *J. Virol.* 65, 3853-3863] and Shibata et al. [Shibata et al. (1995) *J. Virol.* 69, 4453-4462] respectively. Unlike HIV-$1_{LAV\text{-}1b}$, the other parental virus, HIV-1_{SF2}, did not replicate in MDM.

Because the cloned viruses displayed such similar in vitro biologic activities as the uncloned stocks, we sought to examine whether these viruses represented major species in the viral mix. Genomic DNA was isolated and nested PCR was used to amplify the V1-V2 and V3-V5 regions of HIV-1 envelope gene. These hypervariable regions of *env* gene were selected because most viral heterogeneity has been associated with them. Equal amounts of PCR products were mixed, heat denatured, and then reannealed with analogous fragments derived from standards (SF162 subtype B3 and ZM18 subtype C2) or from JC16 and NC7

molecular clones. Electrophoresis of these fragments on polyacrylamide non-denaturing gels revealed differences in migration rates with homoduplexes moving faster in the gel than heteroduplexes. Homoduplexes based on genomic DNA (HIV-1_{JC} or HIV-1_{NC}) resulted in homoduplex (lower) and heteroduplex (upper) bands that had comparable intensities and were close together lane. The HIV-1_{JC}/HIV-1_{JC16} (but not HIV-1_{NC}/HIV-1_{NC7}) heteroduplex band was between the homoduplex and heteroduplex bands obtained with genomic DNA homoduplexes. The distance between homoduplex and heteroduplex bands formed between JC16 and NC7 molecular clones was larger than that between bands formed from homoduplexes derived from genomic DNA. The SF162 (subtype B3) heteroduplex bands migrated at almost the same rate as the ssDNA but the ZM18 (subtype C2) heteroduplex bands had the slowest migration.

An advantage of establishing a model system for AIDS in smaller nonhuman primates is the decreased cost associated with the housing and upkeep of smaller animals. Animal AIDS models include, but are not limited to, chimpanzees, the gibbons, pig-tailed macaques, and rhesus macaques infected with the HIV-1_{JC} or HIV-1_{NC} virus of the invention.

Disclosed herein are an immunodeficient pig-tailed macaque and also an immunodeficient rhesus macaque; both infected with the HIV-1_{JC} virus of the invention isolated from the C499 chimpanzee which exhibited HIV-1-induced AIDS or the HIV-1_{NC} virus isolated from the C455 chimpanzee. The macaques are infected with HIV-1 isolate JC or HIV-1_{NC} and are used as an animal model for human AIDS in essentially the same manner as chimpanzees as described above. The transmission of HIV-1_{JC} or the HIV-1_{NC} virus into a macaque proceeds via injection of an isolated viral suspension or transfection of a biological fluid or tissue specimen from an HIV-1-infected and AIDS-bearing primate. The HIV-1-infected donor specimen is introduced into a recipient by any suitable means, such as intraperitoneal injection, intravenous injection, surgical implantation and combinations thereof. Donor tissue may be introduced as organized tissue (e.g., thymus, lymph node, etc.) or as discrete cells.

Antigenic fragments of the present invention are peptides which contain at least one epitope (antibody binding site) which binds antibodies which bind to at least one HIV-1 isolate of the present invention. The antigenic fragments are preferably capable of inducing an immune response when administered to a nonhuman primate. DNA encoding such antigenic fragments may be used to transform host cells to thereby produce such antigenic fragments.

Antigenic fragments may be identified by a variety of means. A protein from HIV-1_{JC} and/or from HIV-1_{NC}, such as an envelope protein, may be fragmented with a protease, and the fragments tested to determine whether or not various ones react with antiserum against the protein. See, e.g., J. Robinson et al., *Mol. Cell Biochem.* **21**:23-32 (1978). Another technique is to synthesize peptides which are fragments of the entire protein and determine whether the individual fragments are recognized by neutralizing antibodies against the protein. See, e.g., J. Gerin et al., in *Vaccines 85: Molecular and chemical Basis of Resistance to Parasitic, Bacterial and Viral Diseases*, 235-239 (Lerner et al., eds. 1985). Still another method useful for obtaining immunogenic fragments of a protein is by isolation and identification of monoclonal escape mutants. In this strategy, HIV-1 is produced in the presence of a monoclonal antibody to the virus. The only viruses which can grow under these conditions are those with a mutation in the nucleotide sequence which codes for an epitope to which the monoclonal antibody binds. A mutant virus which grows under these conditions is referred to as the "monoclonal escape mutant." The monoclonal escape mutant is then sequenced and the mutant sequence compared with the nucleotide sequence of the HIV-1_{JC} isolate or the HIV-1_{NC} isolate to find the specific location of the mutation. The mutation is located in a region which codes for a protective epitope, or an "immunogenic fragment." See, e.g., J. Lopez et al., *J. Virol.* **64**:927 (1990).

Antigenic preparations of the present invention are useful as reagents in immunoassay diagnostic studies of retroviruses. Immunochemical methods for detecting retroviruses include, for example, immunofluorescence assays or immunoenzymatic assays. Immunofluorescence assays typically involve incubating, for example, serum from the subject to be tested with preparations of the pathogenic virus or fragments thereof. Immune

complexes formed are detected using either direct or indirect methods, for example, the use of antibodies to which fluorescent labels such as rhodamine or fluorescein have been coupled. Immunoenzymatic assays typically involve viral extracts or other antigen-containing compositions bound to a surface. Serum from a subject to be tested for the presence of antibodies directed against one or more antigens is contacted with the surface and, after a period of incubation, unbound substances are washed away. The presence of immune complexes is detected using antibodies labeled with an enzyme such as horseradish peroxidase, alkaline phosphatase, or beta-galactosidase, which is capable of converting a colorless or nearly colorless substrate into a highly colored product, or an enzyme which emits light in the presence of the proper substrate. The amount of product formed is detected visually, spectrophotometrically, or luminometrically and is compared to a similarly treated control. The presence of antibodies in biological fluids may also be detected by agglutination. Viral lysates or antigen compositions are used to coat, for example, latex particles.

Diagnostic tests utilizing the present invention may be carried out in accordance with known techniques. Such techniques provide a method of detecting the presence of HIV-1 by detecting the presence of HIV-1 antibodies. Such methods comprise collecting an antibody-containing biological sample (e.g., blood, blood sera, blood plasma, cerebrospinal fluid, tissue samples) from the subject, contacting the sample with an antigenic preparation of the viral particles of the present invention as given herein, and then detecting the formation of a reaction product between the antibodies in the sample and the antigenic preparation. Any suitable assay format, such as radioimmunoassay (RIA) or enzyme-linked immunosorbent assay (ELISA) may be employed, in accordance with known techniques. See, e.g., *Immunology: Basic Processes*, 162-175 (J. Bellanti, [Ed.] 2d ed., W.B. Saunders Co. (1985).

Also disclosed herein are kits for the detection of HIV infection. Such kits comprise a container containing an antigenic preparation of the viral particles of the present invention, which may be lyophilized. The antigenic preparation may comprise, e.g., the HIV-1 envelope protein (env) or the group antigen (gag) protein of the HIV-1_{JC} and/or HIV-1_{NC} of the invention.

A method of inducing antibodies to HIV-1 in a subject, as disclosed herein, comprises administering to a subject an immunogenic amount of infectious viral particles of the present invention. This method may be used to make polyclonal or monoclonal antibodies, which may be used in diagnostic assays. Suitable subjects include mammals (such as, for example, rats, rabbits, mice, and horses) and primates. The term primates is herein intended to encompass any members of the order Primata (for example, lemurs, mandrills, rhesus monkeys, macaques, and chimpanzees) and to include humans. Suitable subjects include those in which antibodies to HIV may be raised (e.g., rabbit, horse).

In the above methods of inducing antibodies, viral antigenic preparations of the present invention may be combined with any suitable pharmaceutically acceptable carrier (such as sterile, pyrogen-free physiological saline solution, or sterile, pyrogen-free phosphate-buffered saline solution). The viral antigens are included in an effective immunogenic amount. The precise amount to be administered to a given subject is determined by techniques known in the art and will vary depending on the route of administration, the subject and the desired response. Administration to the subject may occur by any suitable route (e.g., by intramuscular injection, subcutaneous injection, intraperitoneal injection, or intravenous injection). The appropriate immunogenic dosage will depend upon the particular subject and the desired outcome. Techniques to determine a particular immunogenic amount of the viral particles of the present invention will be apparent to those of ordinary skill in the art. See, e.g., Johnson et al., *Proc. Natl. Acad. Sci. USA* **89**:2175 (1992). For example, the active agent (viral particles or preparations thereof) may be given in an amount of from 0.05 to 50 μg per kg body weight (e.g., 0.5 or 1.0 μg per kg).

The invention also provides for a variety of different vaccines based on the structures of the HIV-1_{JC} isolate and/or the HIV-1_{NC} of the invention and a method for vaccinating a population against HIV. Examples of active agents used for the preparation of a vaccine of the invention include the live attenuated HIV-1_{JC} and/or HIV-1_{NC} isolate, fixed whole virus, host cells expressing virus antigen, preparations of virus fragments, purified proteins, antigenic fragments of proteins and antigenic peptides which are derivatives of the antigenic fragments. According to the present invention, HIV-1_{JC}- and/or HIV-1_{NC}-derived

compositions or vaccines are useful for preinfection immunization of primates as well as for postinfection (therapeutic) immunization of HIV-infected primates (see Fultz et al. [1989], Lerner et al., eds. Cold Spring Harbor, NY).

Live attenuated HIV isolate JC virus (or HIV-1_{NC} virus) is prepared by serial passage of the virus in tissue culture or genetically altered by recombinant techniques, in accordance with known procedures. Fixed virus is made by contacting live virus (attenuated or unattenuated) to a suitable fixative, such as formalin.

Preparations of viral fragments are made by lysing host cells, such as *E. coli* cells, transformed with a vector encoding an HIV-1 isolate of the present invention (or both) or a portion thereof. The lysate may be used in crude or partially purified form, or a particular viral protein (or antigenic fragment thereof) such as the envelope protein, can be purified to homogeneity and used as an active agent for a vaccine against HIV-1. Host cells such as yeast cells may be transformed with vectors of the present invention capable of expressing HIV-1 proteins, or antigenic fragments thereof, on the surface of the host cells, and the transformed host cells used as an active vaccine agent as such or fixed (e.g., with formalin) and used as an active agent.

Antigenic peptides are selected from the group consisting of antigenic fragments of HIV isolate JC and/or NC proteins, such as the envelope protein, and the antigenic equivalents thereof (i.e., analogs or derivatives). Antigenic peptides may be chemically synthesized or produced by recombinant techniques.

Viral antigenic preparations and cells producing viral antigens and/or fragments thereof may be formulated into immunogenic compositions as neutral or salt forms. Preferably, when cells are used they are of avirulent strains, or the cells are killed before use. Pharmaceutically acceptable salts include but are not limited to the acid addition salts (formed with free amino groups of the peptide) which are formed with inorganic acids, e.g., hydrochloric acid or phosphoric acids; and organic acids, e.g., acetic, oxalic, tartaric, or maleic acid. Salts formed with the free carboxyl groups may also be derived from inorganic

bases, e.g., sodium, potassium, ammonium, calcium, or ferric hydroxides, and organic bases, e.g., isopropylamine, trimethylamine, 2-ethylamino-ethanol, histidine, and procaine.

The term "antigenic equivalents," as used herein, refers to proteins or peptides which bind to an antibody which binds to the protein or peptide with which equivalency is sought to be established. Antibodies which are used to select such antigenic equivalents are referred to as "selection antibodies" herein. Preferred selection antibodies are monoclonal antibodies which bind to HIV isolate JC and/or to HIV-1_{NC}, but not to prior isolates of HIV-1 such as the HIV-1 isolates NDK, LAI and SF2, for example.

One or more amino acids of an antigenic peptide sequence may be replaced by one or more other amino acids which do not affect the antigenicity of that sequence. Such changes can be guided by known similarities between amino acids in physical features such as charge density, hydrophobicity/hydrophilicity, size and configuration, For example, threonine and serine can be interchanged, or aspartic acid and glutamic acid, or leucine and isoleucine, and the like.

Antigenic equivalents may be formed by modifying reactive groups within a natural sequence or modifying the N-terminal amino and/or C-terminal carboxyl group. Such equivalents include salts formed with acids and/or bases, particularly physiologically acceptable inorganic and organic acids and bases. Other equivalents include modified carboxyl and/or amino groups on the synthetic peptide to produce esters or amides, or amino acid protecting groups such as N-t-butoxycarbonyl. Preferred modifications are those which provide a more stable, active peptide which will be less prone to enzymatic degradation *in vivo*.

In a particular embodiment of the invention, polyclonal and/or monoclonal antibodies capable of specifically binding to a particular epitope of at least one HIV-1 isolate of the invention are provided. The term *antibody* is used to refer both to a homogenous molecular entity (monoclonal antibody) and a mixture (such as a serum product) made up of a plurality of different molecular entities (polyclonal antibody). Monoclonal or polyclonal antibodies,

and preferably monoclonal, specifically reacting with a particular epitope of interest can be made by methods known in the art. See, e.g., Harlow and Lane (1988) *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratories; Goding (1986) *Monoclonal Antibodies: Principles and Practice*, 2d ed., Academic Press, New York. Also, recombinant immunoglobulins may be produced by methods known in the art, including but not limited to the methods described in U.S. Patent No. 4,816,567, incorporated by reference herein. Monoclonal antibodies with affinities of 10^8 M^{-1}, preferably 10^9 to 10^{10} or more are preferred.

For use as a vaccine, immunogenic compositions may be formulated by any of the means known in the art. Such vaccines are typically prepared as injectables, either as liquid solutions or suspensions. Solid forms suitable for solution in, or suspension in, liquid prior to injection may also be prepared. The preparation may also, for example, be emulsified, or the protein encapsulated in liposomes. Such vaccines may be administered to the subject by any suitable means, for example, by intramuscular injection, by subcutaneous injection, by intravenous injection, by intraperitoneal injection, by oral injection, and by nasal spray.

The vaccine or other immunogenic composition may be given in a single dose or multiple dose schedule. A multiple dose schedule is one in which a primary course of vaccination may include 1 to 10 or more separate doses, followed by other doses administered at subsequent time intervals as required to maintain and or reinforce the immune response, e.g., at 1 to 4 months for a second dose and, if needed, a subsequent dose(s) after several months.

The immunogenic peptide antigen compositions are administered in a manner compatible with the dosage formulation and in such amount as will be prophylactically and/or therapeutically effective. The quantity to be administered, which is generally in the range of about 100 to 1,000 μg of protein per dose, more generally in the range of about 5 to 500 μg of protein per dose, depends on the subject to be treated, the capacity of the individual's immune system to synthesize antibodies, and the degree of protection desired. Precise amounts of the active ingredient required to be administered may depend on the judgment of the physician

and may be peculiar to each individual, but such a determination is within the skill of such a practitioner.

Vaccine formulations of the present invention comprise the active agent mixed with excipients or carriers which are pharmaceutically acceptable and compatible with the active ingredient. Suitable excipients include but are not limited to water, saline, dextrose, glycerol, ethanol, or the like and combinations thereof. The concentration of the immunogenic polypeptide in injectable formulations is usually in the range of 0.2 to 5 mg/ml.

In addition, if desired, the vaccines may contain minor amounts of auxiliary substances such as wetting or emulsifying agents, pH buffering agents, and/or adjuvants which enhance the effectiveness of the vaccine. Examples of adjuvants which may be effective include but are not limited to: aluminum hydroxide; aluminum phosphate; plant and animal oils; synthetic polymers; e.g., N-acetyl-muramyl-L-threonyl-D-isoglutamine (thr-MDP); N-acetyl-nor-muramyl-L-alanyl-D-isoglutamine (CGP 11637, referred to as nor-MDP); N-acetylmuramyl-L-alanyl-D-isoglutaminyl-L-alanine-2-(1'-2'-dipalmitoyl-sn-glycero-3-hydroxyphosphoryloxy)-ethylamine (CGP 19835A, referred to as MTP-PE); and RIBI, which contains three components extracted from bacteria, monophosphoryl lipid A, trehalose dimycolate; cell wall skeleton (MPL+TDM+CWS) in a 2% squaline/Tween 80 emulsion; etc.

In addition, the vaccine formulations may also contain one or more stabilizer, for example, carbohydrates such as sorbitol, mannitol, starch, sucrose, dextrin, and glucose, proteins such as albumin or casein, and buffers such as alkaline metal phosphates and the like.

In view of the similarities in protein molecules making up different isolates of HIV-1, the skilled artisan understands that an antibody, particularly a monoclonal antibody, which is specific for a particular epitope directed to a particular protein of the HIV-1_{JC} isolate and/or the HIV-1_{NC} isolate, can be used to screen for other HIV-1 isolates having similar epitopes recognized by that (monoclonal) antibody.

Antibodies generated against specific epitopes of the HIV-1_{JC} of the invention are useful, for example, as probes for screening DNA expression libraries or for detecting the presence of HIV-1 strains in a test sample. Antigens can be synthesized and conjugated to a suitable carrier protein (e.g., bovine serum albumin or keyhole limpet hemocyanin) for use in vaccines or in raising specific antibodies. Frequently, the polypeptides and antibodies will be labeled by joining, either covalently or noncovalently, a substance which provides a detectable signal. Suitable labels include but are not limited to radionuclides, enzymes, substrates, cofactors, inhibitors, fluorescent agents, chemiluminescent agents, magnetic particles and the like. United States Patents describing the use of such labels include but are not limited to Nos. 3,817,837; 3,850,752; 3,939,350; 3,996,345; 4,277,437; 4,275,149; and 4,355,241.

Immunogenic carriers may be used to enhance the immunogenicity of an active agent. Such carriers include but are not limited to proteins and polysaccharides, liposomes, and bacterial cells and membranes. Protein carriers may be joined to the capsular polysaccharide molecules to form fusion proteins by recombinant or synthetic means or by chemical coupling. Useful carriers and means of coupling such carriers to polypeptide antigens are known in the art. The art knows how to administer immunogenic compositions so as to generate protective immunity where immunity is most helpful.

Compositions and immunogenic preparations including vaccine compositions comprising substantially purified antigens derived from an HIV-1 isolate JC and/or NC and a suitable carrier therefor are provided. Immunogenic compositions are those which result in specific antibody production when injected into a human or an animal. Such immunogenic compositions are useful, for example, in immunizing primates against infection by HIV-1 strains. The immunogenic preparations comprise an immunogenic amount of, as specifically exemplified, at least one antigenic determinant derived from the HIV-1_{JC} isolate of the invention and a suitable carrier. Alternatively, the immunogenic composition can comprise host cells harboring an antigenic agent from the specifically exemplified HIV-1_{JC} strain and a suitable carrier. It is understood by one of ordinary skill in the art that a functionally equivalent, recombinant mutant of HIV-1_{JC} and or HIV-1_{NC} can be produced by the

introduction of the cloned DNA containing the insertion mutations responsible for a desired characteristic. It is also within the scope of the present invention and readily within the grasp of the ordinary skilled artisan to generate other types of genetically stable mutations in the structural or enzyme genes of HIV-1. Such immunogenic compositions (or vaccines) are useful, for example, in immunizing an animal, especially humans, against AIDS and related diseases resulting from infection by HIV-1 species. Such immunogenic compositions can also elicit the production of antibodies which will cross react with proteins of other HIV-1 and HIV-2 strains expressing epitopes in common with those of the starting HIV-1_{JC} isolate. It is understood that where whole cells are formulated into the immunogenic composition, the cells are preferably inactivated, especially if the cells are of a virulent strain. Such immunogenic compositions may comprise one or more protein or the immunogenic cellular component. By "immunogenic amount" is meant an amount capable of eliciting the production of antibodies directed against an antigenic agent of HIV-1_{JC} in an animal or human to which the vaccine or immunogenic composition has been administered.

The nucleotide sequence of the HIV-1_{JC} isolate or the HIV-1_{NC} isolate can be used to generate hybridization probes which specifically bind to HIV-1_{JC} genetic material, or to DNA of HIV-1 isolates having the identifying characteristics of the HIV-1 isolates JC or NC, to determine the presence of such HIV-1 in primates. The hybridization probe may be selected so that it does not bind to other known HIV-1 isolates such as NDK, LAI, SF2, etc. The hybridization probes can be cDNA fragments or polynucleotides and may be labeled with a detectable group, as is well-known in the art. Pairs of probes can serve as PCR primers for synthesis and amplification processes in accordance with the description, for example in U.S. Patent Nos. 4,683,202 and 4,683,195.

In specific embodiments, probes of the invention comprise DNA sequences of HIV-1_{JC} or sequences encoding antigenic fragments thereof or sequences having identity thereto. In particular are provided probes having a DNA sequence as set forth in SEQ ID NO:1 or a sequence having identity thereto.

The production of DNA, vectors, transformed host cells, HIV-1 virus, proteins, and protein fragments of the present invention by genetic engineering techniques can be carried out in accordance with methods known in the art. See, e.g., U.S. Pat. No. 4,761,371, U.S. Pat. No. 4,877,729, U.S. Pat. No. 4,912,038, and U.S. Pat. No. 4,879,224, among others.

A nucleotide sequence (polynucleotide) or fragment thereof is substantially homologous (or substantially similar) to another polynucleotide if, when optimally aligned (with appropriate nucleotide insertions or deletions) with another polynucleotide, there is nucleotide sequence identity for approximately 80% of the nucleotide bases, usually approximately 90%, more preferably about 95% to 100% of the nucleotide bases. Gaps introduced to optimize alignments are treated as mismatches.

Alternatively, substantial homology (or similarity) exists when a polynucleotide or fragment thereof will hybridize to another polynucleotide under selective or stringent hybridization conditions. Selectivity of hybridization exists under stringent hybridization conditions which allow one to distinguish the target polynucleotide of interest from other polynucleotides. Typically, selective hybridization will occur when there is approximately 75% similarity over a stretch of about 14 nucleotides, preferably approximately 80% similarity, more preferably approximately 85% similarity, and most preferably approximately 90% similarity. See Kanehisa (1984) *Nucl. Acids Res.*, **12**:203-213. The length of homology comparison, as described, may be over longer stretches, and in certain embodiments will often be over a stretch of about 17 to 20 nucleotides, preferably 21 to 25 nucleotides, more preferably 26 to 35 nucleotides, and more preferably about 36 or more nucleotides.

The hybridization of polynucleotides is affected by such conditions as salt concentration, temperature, or organic solvents, in addition to the base composition, length of the complementary strands, and the number of nucleotide base mismatches between the hybridizing polynucleotides, as will be readily appreciated by those skilled in the art. Stringent temperature conditions will generally include temperatures in excess of 30°C, typically in excess of 37°C, and preferably in excess of 45°C. Stringent salt conditions will ordinarily be less than 1 M, typically less than 500 mM, and preferably less than 200 mM.

However, the combination of parameters is much more important than the measure of any single parameter (Wetmur and Davidson [1968] *J. Mol. Biol.* 31:349-370).

An isolated or substantially pure polynucleotide is a polynucleotide which is substantially separated from other polynucleotide sequences which naturally accompany a native sequence. The term embraces a polynucleotide sequence which has been removed from its naturally occurring environment, and includes recombinant or cloned DNA isolates, chemically synthesized analogues and analogues biologically synthesized by heterologous systems.

A polynucleotide is said to encode a polypeptide if, in its native state or when manipulated by methods known to those skilled in the art, it can be transcribed and/or translated to produce the polypeptide of a fragment thereof. The anti-sense strand of such a polynucleotide is also said to encode the sequence.

Vectors are replicable DNA constructs used to either amplify or express DNA of the present invention. An expression vector is a replicable DNA construct in which DNA of the present invention is operably linked to control sequences capable of expressing that DNA in a suitable host. Generally, control sequences include a transcriptional promoter, an optional operator sequence to control transcription, a sequence encoding suitable mRNA ribosomal binding sites, and sequences which control the termination of transcription and translation. Suitable vectors include plasmids, viruses (e.g., vaccinia virus, adenovirus, baculovirus, cytomegalovirus) phage, and integratable DNA fragments (i.e., fragments integratable into the host genome by recombination).

DNA regions are operably linked or operably associated when they are functionally related to each other. For example, a promoter is operably linked to a coding sequence if it controls the transcription of the sequence; or a ribosome binding site is operably linked to a coding sequence if it is positioned so as to permit translation.

Transformed host cells are cells which have been transformed or transfected with vectors as described above. Transformed host cells ordinarily express the DNA of the present invention. Suitable host cells include prokaryote, yeast or higher eukaryotic cells such as mammalian cells and insect cells.

Prokaryote host cells include gram negative or gram positive organisms, for example *Escherichia coli (E. coli)* or Bacilli. Exemplary host cells are *E. coli* W3110 (ATCC 27,325), *E. coli* B, *E. coli* X1776 (ATCC 31,537), *E. coli* 294 (ATCC 31,446). A broad variety of suitable prokaryotic and microbial vectors are available. *E. coli* is typically transformed using pBR322. Promoters most commonly used in recombinant microbial expression vectors include the β-lactamase (penicillinase) and lactose promoter systems (Change et al., *Nature* **275**:615 [1978]; and Goeddel et al., *Nature* **281**:544 [1979], a tryptophan (trp) promoter system (Goeddel et al. [1980] Nuc. Acids Res. **8**:4057, and EPO App. Publ. No. 36,766) and the tac promoter (H. De Boer et al., *Proc. Natl. Acad. Sci. USA* **80**:21 [1983]). The promoter and Shine-Dalgarno sequence are operably linked to the DNA of the invention, i.e., they are positioned so as to promote transcription of messenger RNA from the DNA.

Eukaryotic microbes such as yeast cultures may also be transformed with vectors of the present invention. See, e.g., U.S. Pat. No. 4,745,057. *Saccharomyces cerevisiae* is the most commonly used yeast, although other yeast may also be used. Host cells such as insect cells (e.g., cultured *Spodoptera frugiperda* cells) and expression vectors such as the baculovirus expression vector may be employed in carrying out the present invention, as described in U.S. Pat. Nos. 4,745,051 and 4,879,236 to Smith et al.

Examples of useful mammalian host cell lines are VERO and HeLa cells, Chinese hamster ovary (CHO) cell lines, and WI138, BHK, COS-7, CV, and MDCK cell lines. The transcriptional and translational control sequences in expression vectors to be used in transforming vertebrate cells are often provided by viral sources. For example, commonly used promoters are derived from polyoma, Adenovirus 2, and Simian Virus 40 (SV40). See, e.g., U.S. Pat. No. 4,599,308. An origin of replication may be provided either by construction of the vector to include an exogenous origin, such as may be derived from SV40 or other viral

(e.g., Polyoma, Adenovirus, VSV, or BPV) source, or may be provided by the host cell chromosomal replication mechanism. If the vector is integrated into the host cell chromosome, the latter is often sufficient. Rather than using vectors which contain viral origins of replication, one can transform mammalian cells by the method of cotransformation with a selectable marker and DNA of the present invention, as described in U.S. Pat. No. 4,399,216.

Except as noted hereafter, standard techniques for cloning, DNA isolation, amplification and purification, for enzymatic reactions involving DNA ligase, DNA polymerase, restriction endonucleases and the like, and various separation techniques are those known and commonly employed by those skilled in the art. A number of standard techniques are described in Sambrook et al. (1989) *Molecular Cloning*, Second Edition, Cold Spring Harbor Laboratory, Plainview New York; Maniatis et al. (1982) *Molecular Cloning*, Cold Spring Harbor Laboratory, Plainview, New York; Wu (ed.) (1993) *Meth. Enzymol.* **218**:Part I; Wu (ed.) (1979) *Meth. Enzymol.* **68**; Wu et al. (eds.) (1983) *Meth. Enzymol.* **100** and **101**; Grossman and Moldave (eds.) *Meth. Enzymol.* **65**; Miller (ed.) (1972) *Experiments in Molecular Genetics*, Cold Spring Harbor Laboratory, Plainview, New York; Old and Primrose (1981) *Principles of Gene Manipulation*, University of California Press, Berkely; Schleif and Wensink (1982) *Practical Methods in Molecular Biology*; Glover (ed.) (1985) *DNA Cloning* Vol. I and II, IRL Press, Oxford, UK; Hames and Higgins (eds.) (1985) *Nucleic Acid Hybridization*, IRL Press, Oxford, UK; and Setlow and Hollaender (1979) *Genetic Engineering: Principles and Methods*, Vols. 1-4, Plenum Press, New York. Abbreviations and nomenclature, where employed, are deemed standard in the field and commonly used in professional journals such as those cited herein.

All publications, patent applications and patents cited herein are incorporated by reference to the extent that they are not inconsistent with the present disclosure.

The foregoing discussion and the following examples are provided for illustrative purposes, and they are not intended to limit the scope of the invention as claimed herein. Modifications and variations which may occur to one of ordinary skill in the art are within the

intended scope of this invention. The present invention is further described in the non-limiting examples set forth below.

EXAMPLES

Example 1. Animal Subjects

All animals (e.g., chimpanzees) were maintained in accordance with the guidelines established by the Animal Welfare Act and the NIH guide for care and use of laboratory animals. The Yerkes Center is fully accredited by the American Association for the Accreditation of Laboratory Animal Care (AAALAC).

A cohort of 12 chimpanzees was inoculated with several strains of HIV-1 at the Yerkes Center. A member of this cohort, chimpanzee 499 was inoculated on three different occasions with three different HIV-1 isolates: HIV-1_{SF2} in 1985, HIV-1_{LAI} in 1986 and HIV-1_{NDK} in 1987 (See Fultz et al. [1991] J. Infect. Dis. **163**:441-447 and Novembre et al. [1997] J. Virol. **71**:4086-4091).

It is preferred that a chimpanzee be bred in captivity, be seropositive for Epstein-Barr virus and cytomegalovirus and be seronegative and PCR negative for HIV prior to being used as an animal model.

Example 2. Isolation and Purification of HIV-1_{JC} Isolate

The original source of the HIV-1_{JC} isolate was the chimpanzee C499. HIV was easily isolated from the peripheral blood mononuclear cells (PBMC) of this animal. Cocultivation of PBMC derived from C499 with uninfected chimpanzee PBMC (cPBMC) resulted in the isolation of a virus (HIV-1_{JC}) which induced syncytium formation in chimpanzee cells. The nucleotide sequence of this virus was distinguished from the isolates used for the initial inoculations and from other known HIV-1 isolates.

The HIV-1_{JC} isolate was deposited with the AIDS Reagent Program, McKesson Bioservices, 685 Lofstrand Lane, Rockville, MD 20850 USA, a division of the NIH AIDS

Research and Reference Reagent Program, in May 1997 and has been assigned Catalog Number 3523. A substantially full-length infectious molecular clone has a nucleotide sequence as given in SEQ ID NO:11.

In a specific embodiment of the invention, probes of the invention comprise DNA sequences as set forth in SEQ ID NO:11 and/or sequences of at least 15 contiguous nucleotides derived therefrom or sequences complementary thereto.

The original source of the HIV-1_{JC} isolate was the chimpanzee C455. HIV was easily isolated from the peripheral blood mononuclear cells (PBMC) of this animal. Cocultivation of PBMC derived from C455 with uninfected chimpanzee PBMC (cPBMC) resulted in the isolation of a virus (HIV-1_{NC}) which induced syncytium formation in chimpanzee cells. The nucleotide sequence of this virus is distinguished from other known HIV-1 isolates. A substantially full-length infectious molecular clone has a nucleotide sequence as given in SEQ ID NO:12.

In a specific embodiment of the invention, probes of the invention comprise DNA sequences as set forth in SEQ ID NO:12 and/or sequences, of at least 15 contiguous nucleotides derived therefrom or sequences complementary thereto.

Example 3. Virus and Viral Antigens

The HIV-1 isolates used to inoculate C499 included LAV (lymphadenopathy-associated virus)-1_{BRU} (Barre-Sinoussi [1983] Science **220**:868-871), SF2 (previously designated ARV [AIDS-related virus]-2) (Levy et al. [1984] Science **225**:840-842), and NDK, a highly cytopathic HIV-1 of African origin (Spire et al. [1989] Gene **81**:275-284). Inoculations of virus were done intravenously with 1-ml aliquots of undiluted or diluted virus stocks. In January 1988, 33 months after inoculation HIV-1_{SF2}, C499 was part of a study to assess the effects of therapeutic vaccination on immunity and viral status in HIV-1-infected chimpanzees. C499 was given two intramuscular injections, 4 weeks apart, of 500 µg of recombinant SF2 p53 *gag*, produced in yeast and formulated with 100 µg of muramyl

tripeptide (Ciba-Geigy, Summit, NJ) in 4% squalene and 0.008% Tween 80 (Fultz et al. [1989] in Vaccines 89, *supra*).

The HIV-1 isolates used to inoculate C455 was JC, deposited with the AIDS Reagent Program, Catalog Number 3523. Inoculations of virus were done intravenously with 1-ml aliquots of undiluted or diluted virus stock. C534 was given two intramuscular injections, 4 weeks apart, of 500 μg of recombinant SF2 p53 *gag*, produced in yeast and formulated with 100 μg of muramyl tripeptide (Ciba-Geigy, Summit, NJ) in 4% squalene and 0.008% Tween 80 (Fultz et al. [1989] in Vaccines 89, *supra*).

Example 4. Serologic Assays

Serum samples were tested by indirect ELISA for antibodies to specific HIV-1 proteins using a series of recombinant antigens. These antigens have been described (Fultz et al. [1989] in *Vaccines* 89, *supra)* and included p25 *gag* and p53 *gag*, produced in yeast; p31 *pol*; and *env*2-3(SF2) and *env*2-3 (IIIB), nonglycosylated polypeptides corresponding to full-length gpl20. Antibody titers to whole-virus preparations were determined with an HIV enzyme immunoassay kit (EIA; Genetic Systems, Seattle, WA). Neutralization assays were performed as described (Fultz et al. [1986] Proc. Natl. Acad. Sci. **83**:5286-5290); titers were based on >80% inhibition of reverse transcriptase (RT) activity after preincubation of virus with serum and infection of normal human PBMC. Antibodies cross-reactive with histone H2B were identified by immunoblot using purified calf thymus histones as described previously (Strickler et al. [1987] Nature **327**:710-713).

Example 5. Virus Assays

Assays to detect virus in plasma samples or to recover virus from chimpanzee PBMC were performed with normal human PBMC as indicator cells, unless otherwise indicated, and have been described (Fultz et al. [1986] J. Virol. **58**:116-124). Medium was RPMI 1640 with 10% fetal bovine serum, 8 units of recombinant interleukin-2 (IL-2)/ml, glutamine, and antibiotics (RPMI-IL-2). Infectious PBMC were quantified by serial 1:5 limiting dilution and cocultivation with human PBMC (Fultz et al. [1986], *supra*). HIV-1 was detected in all assays by the presence of RT activity in cell-free culture supernatants. To detect HIV antigen in

serum or plasma samples, a commercially available HIV-1 antigen capture kit (Coulter, Hialeah, FL) was used.

To assess the influence of $CD8^+$ lymphocytes on viral replication and recovery, after 3 days of stimulation with concanavalin A (ConA 10 μg/ml) in RPMI 1640, PBMC were washed once; 10^7 cells from C499 were placed in fresh RPMI-IL-2, cultured without indicator cells, and monitored for HIV-1 production. In other experiments, the overnight incubation of C499's PBMC to remove adherent cells, $CD4^+$- and $CD8^+$-enriched populations were obtained by panning with monoclonal antibodies specific for the $CD8^+$ antigen or with $CD8^+$-coated magnetic beads (Dynabeads; Robbins Scientific, Mountain View, CA). After removal of the $CD8^{-[??]}$ cells, $CD8^+$ lymphocytes were recovered from the plates or magnetic beads by additional overnight incubation and repeated washes with medium. After stimulation with ConA, cultures were established either with $CD4^+$-enriched cells only or with $CD4^+$- and $CD8^+$-enriched cells in various ratios. Percentages of $CD4^+$ and $CD8^+$ cells in enriched populations were determined by analysis with FACScan (Becton-Dickinson, Mountain View, CA).

Replication kinetics in PBMC and macrophages were tested as follows. 1×10^7 freshly isolated or Con-A stimulated PBMC from HIV-1 negative chimpanzee were infected overnight (at 37°C) with 20 ng of the indicated virus (p24 antigen concentration). The cells were centrifuged at 1000 rpm for 10 min, resuspended in 10 ml complete RPMI medium containing 10% FBS and IL-2 (IL-2 medium) and were incubated at 37°C. Samples of supernatants (1 ml) were harvested on days 3, 7, 10, 14 and 17 post infection. IL-2 medium was added to the cultures following the sampling to maintain the original volume. Supernatants were used in RT assays to determine the relative amounts of virus produced.

For replication in macrophages, cPBMC were resuspended in macrophage media (6×10^6/well in RPMI 1640 containing 15% human serum [AB^+], 1% HEPES, 0.008 ng/ml GM-CSF, 0.03 ng/ml M-CSF, 1% antibiotic-antimycotic solution (Sigma, St. Louis, MO) and seeded in a 24-well plate and incubated at 37°C for 4 hrs. The cells were mixed by pipetting up and down before the incubation was continued for 4 days. Non-adherent cells were

removed by gently washing the wells. Fresh medium (2 ml/well) was added and the cells were cultured for an additional 3 days to allow full macrophage differentiation. Infections were initiated by adding 10 ng of virus (p24) to the cells in 500 µl media and adsorbed overnight. The inoculum was removed and the cells were washed twice before fresh macrophage medium (2 ml) was added. On days 7 and 14 post infection, aliquots of 0.5 ml were taken for determination of p24 antigen levels using the HIV-1 p24 Antigen kit (Coulter Corp, Miami, FL) according to the manufacturer's instructions.

Example 6. Cell-mediated Immune Responses

Lymphocyte proliferative responses to mitogens were tested by incorporation of [^{3}H]thymidine into the DNA of PBMC seeded in triplicate at 5 x 10^4 PBMC per well into 96-well plates. PBMC were incubated for four days with different concentrations of phytohemagglutinin (PHA) or ConA, pulsed overnight with 1 µCi of [^{3}H]thymidine per well, and harvested with a cell harvester (Skatron, Sterling, VA); counts per minute (cpm) incorporated were determined in a β counter.

Cytotoxic T lymphocyte (CTL) responses against HIV-1 *env*- and *gag*-encoded antigens were assayed using fresh PBMC from C499 as effector cells and autologous Epstein Barr virus-transformed B cells as target cells. Target cells were infected at MOI=10 with recombinant vaccinia viruses expressing HIV-1 gpl 20 *env* of p24 *gag* or with wild-type vaccinia as control. Infected targets were labeled with 200 µCi of [^{51}Cr] sodium chromate (Du Pont, Boston) for 2 hours and washed, and 10^5 target cells (T) were mixed with effector cells (E) at various ratios in 96-well plates. The plates ere incubated for 6 hours at 37°C in 5% CO_2, then 0.1 ml of medium was removed from each well and counted in a γ counter. Maximum release (max) was determined by adding 1% Triton X to labeled target cells, and spontaneous release (spon), by adding only medium. Percentage of specific lysis was determined: [cpm(experimental) - cpm (spon)]/[cpm(max) - cpm(spon)] x 100.

Example 7. Nonhuman Primate Model Systems for AIDS

Nonhuman primates infected with HIV-1 isolate JC or infected with HIV-1_{NC} are useful as a model system for the study of AIDS. Chimpanzees and other monkey species

used for this purpose are preferably specific pathogen-free animals, which are available from primate centers, e.g., the Yerkes Regional Primate Research Center, Emory University, Atlanta, GA.

Infected primates are preferably maintained as a single colony of two or more animals, all inoculated with HIV-1 isolate JC or a colony of two or more animals, all inoculated with HIV-1_{NC}. A colony may be maintained in a single room with each primate housed in an appropriate cage, in accordance with standard practices for the maintenance of animals established by the Animal Welfare Act and the NIH guide for care and use of laboratory animals.

The primates are infected with the HIV-1_{JC} virus or the HIV-1_{NC} virus by any suitable means, such as intraperitoneal, intravenous or subcutaneous injection with a solution containing HIV-1 isolate JC or HIV-1_{NC}. The solution may also be a body fluid or tissue (e.g., blood) from a previously infected primate, a blood fraction containing peripheral blood mononuclear cells from a previously infected primate, a pharmaceutically acceptable carrier such as saline solution containing HIV-1 isolate JC or HIV-1 isolate NC, etc.

Nonhuman primates infected with HIV-$_{JC}$ or HIV-1_{NC} are particularly useful as a model system for AIDS because of the concomitant decrease in $CD4^+$ cells and increase in HIV-1 loads in plasma. The development of AIDS has not been previously documented in any nonhuman primate species. When used as a model system, a primate(s) infected with HIV-1_{JC} or HIV-1_{NC} virus is subjected to a treatment useful in combating AIDS in humans and thereafter the progress of the infection and related diseases is monitored. A control (placebo) group of HIV-1_{JC} or HIV-1_{NC}-infected animals is left untreated for comparative purposes. A slowing in the progression of the development of AIDS in infected animals indicates that the treatment may be useful for combating AIDS in humans and additional screening and toxicological testing is prescribed. Such treatment includes but is not limited to a vaccine, a drug (e.g., an antiretroviral compound) or a drug combination (e.g., antiviral nucleosides such as AZT, DDI, etc.), a peptide, a protein, etc.) or a vaccine/drug combination, etc. After treatment initiation, the progress of the disease is monitored by any suitable

parameters including, but not limited to, (a) decline in $CD4^+$ cell levels, (b) increase in viral loads in plasma, (c) presence of HIV-1_{JC} virus or HIV-1_{NC} virus, (d) weight loss, (e) general appearance, and other symptoms characteristic of AIDS.

Example 8. Virus Cloning and Sequence Analysis

cPBMC infected with either HIV-1_{JC} or HIV-1_{NC} were used for isolation of DNA using the Puregene Kit (Gentra systems, Minneapolis, MN) as directed by the manufacturer. The strategy used for PCR amplification and primer location on the HIV-1 viral genome is shown in Figure 3. PCR primers MSF12 (5'-AAA TCT CTA GCA GTG GCG CCC GAA CAG-3', (SEQ ID NO:18); HIV-1_{LAV} nt 169 to 195) and MSR5 (5'-GCA CTC AAG GCA AGC TTT ATT GAG GCT-3', (SEQ ID NO:19); HIV-1_{LAV} nt 9225 to 9198) [Salminen et al. (1995) *Virology* 213, 80-86] were used to amplify a 9056 bp product from PBMC genomic DNA prepared from HIV-1_{JC}-infected cells. Another PCR primer pair 527 (5'-CAC ACA CAA GGC TAC TTC CCT GAT TGG CAG A-3', SEQ ID NO:20, HIV-1_{LAV} nt 5302 to 5274) was used to amplify 5' LTR-containing fragments (5699 bp) from the same DNA source. For the HIV-1_{NC} viral genome, PCR primer pairs 527-528 and 529 (5'-ATG GAA CAA GCC CCA GAA GAC CAA GGG CCA CAG-3', SEQ ID NO:21, HIV-1_{LAV} nt 5141 to 5173) and 530 (5'-GGT CTG AGG GAT CTC TAG TTA CCA GAG TCA C-3', SEQ ID NO:22, HIV-1_{LAV} nt 151 to 121) were used to generate the 5'-half (5699 bp) and 3'-half (4142 bp) PCR products respectively from HIV-1_{NC} genomic PBMC DNA. Primers were synthesized on an Applied Biosystems 392 DNA synthesizer (Applied Biosystems, Foster City, CA). Briefly, PCR was performed using the reagents from the Expand Long Template kit (Boehringer Mannheim, Indianapolis, IN) and 200 ng of DNA template, according to the manufacturer's instructions. After an initial DNA denaturation of 94°C for 2 min, the PCR consisted of 10 cycles of 94°C for 15s, 61°C for 30s, 68°C for 8 min followed by 20 cycles of 94°C for 15s, 61°C for 30s, and 68°C for 8 min with a 5 second addition to each extension. The samples were incubated at 72°C for 30 min after the last cycle and then cooled to 4°C. Results of PCR reactions were evaluated on 0.9% agarose gels. PCR products representing the correct sized fragments were isolated from agarose and were directly cloned into the pCR II plasmid and amplified in *Escherichia coli* bacteria (TA cloning kit, Invitrogen Corp., San Diego, CA) according to the manufacturer's protocol.

Single bacterial colonies containing plasmids with inserts of the correct size were grown at 30°C overnight and plasmid DNA prepared by the alkaline lysis method.

The strategy for preparing full-length molecular clones is illustrated in Figure 3. Several restriction enzymes were used to generate restriction maps for the positive clones. The *Apa* I fragment (1947 bp) from the 5' half PCR product was gel-purified and subcloned into the large fragment (7675 bp) of pJC using standard cloning procedures [Sambrook et al. (1989) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY] to generate plasmid pHIV-1_{JC16} (Fig. 1). A chimeric plasmid (pHIV-1_{NCJC}) was generated by subcloning the PCR amplified 5' half of HIV-1_{NC} into pHIV-1_{JC16}. For HIV-1_{NC} full-length clone, the *Nco* I-*Xho* I *env*-containing fragment from the 3' half PCR product was gel purified and subcloned into the *Nco* I-*Xho* I large fragment of plasmid pHIV-1_{NCJC} containing the 5' half of HIV-1_{NC}. Multiple restriction enzymes were used for analysis of both viral DNAs to confirm the full length clones.

5 x 10^6 CEMx174 cells in T-25 flasks were transfected with 2 µg of either pHIV-1_{JC16} or NC7 DNA in transfection buffer (25 mM Tris-HCl)(pH 7.5, 140 mM NaCl, 5 mM Kcl, 0.7 mM $K_2HPO_43H_2O$) containing 4 µl of DEAE-dextran (60 mg/ml) for 20 minutes at room temperature. Five milliliters of complete medium (RPMI 1640 supplemented with 10% fetal bovine serum (FBS) and 2 mM L-glutamine) was added to stop the reaction followed by centrifugation of cells at 1000 rpm for 10 minutes. The cells were washed twice in 10 ml complete medium before they were resuspended in 10 ml complete medium, transferred to a T-25 cm^2 flask and incubated at 37°C (5% CO_2). The cells were checked daily for cytopathic effects (syncytia formation) and aliquots of cultures were tested for the presence of reverse transcriptase (RT) activity using standard assay methods. For 293 cell lines, 2 x 10^5 cells in 6-well plates were transfected with 2 µg of viral DNA using lipofectin (Life Technologies, Gaithersburg, MD) or DOTAP (Boehringer Mannheim, Indianapolis, IN) according to the manufacturers' instructions. After 24 hours, the transfected cells were overlaid with 2 x 10^6/well of uninfected cPBMC previously stimulated with concanavalin A (Con-A) for 4 days. After an additional 2 day incubation, the non-adherent cell population (cPBMC) were transferred to a T-25 flask and additional stimulated cPBMC added for virus amplification.

Culture supernatants were assayed for RT activity and the cells were observed daily for development of syncytia. Cell-free stocks of molecularly cloned viruses were prepared at peak RT activity, aliquoted and stored under liquid nitrogen.

Primers for sequencing were constructed from conserved regions of aligned sequences of HIV-1_{LAI} and HIV-1_{SF2} and were synthesized on an applied Biosystems 392 DNA synthesizer. The DNA sequence of each full-length clone virus was determined by the dideoxy-chain termination method using the sequenase system (Amersham Life Sciences, Arlington Heights, IL) and ^{35}S-dATP. Nucleotide sequence alignments were performed with the Intelligenetics Suite of programs (Intelligenetics, Beaverton, OR) whild the phylogenetic analysis of amino acid sequence was done with CLUSTAL method (DNASTAR, Inc., Madison, WI).

Example 9. Heteroduplex Mobility Assays.

The nucleotide sequences for HIV-1_{JC16} and HIV-1_{NC7} have been assigned Genbank accession numbers AF049494 and AF049495 respectively.

The heteroduplex mobility assay kit (NIH AIDS Research and Reference Reagent Program) based on the method described by Delwart et al. (1993) *Science* 262, 1257-1261 was used. Briefly, equal amounts (5 µl each) of second-round PCR products (V1-V2 and V3-V5) from infected cPBMC genomic DNA were mixed with the reference PCR products to obtain heteroduplexes. After adding 1.1 µl of 10x annealing buffer (1 M NaCl, 100 mM Tris [pH 7.8], 20 mM EDTA), the mixed DNAs were denatured at 94°C for 2 min and then reannealed by rapidly cooling in ice. Three µl of loading dye (25% Ficoll, 1% Orange G) was added to the cooled DNA mixture and the samples loaded onto 5% polyacrylamide gel in 1x TBE (88 mM Tris-borate, 89 mM boric acid, 2 mM EDTA) buffer and electrophoresed at a constant voltage of 250 V for 2.5 hrs. The gels were stained with ethidium bromide and visualized under ultra-violet (UV) light.

TABLE 1

Platelet counts, lymphocyte subset levels, and plasma HIV-1 loads in C499

Date (mo/day/yr)	Platelet count (cells/μl)	Total lymphocyte count (cells μl)[a]	Absolute CD4+ cell count (cells/μl)	Absolute CD8+ cell count (cells/μl)	Plasma virus level (RNA equivalents/ml)[b]
10/04/88	63,000	3,136	550	1,020	$<10^4$
06/12/89	122,000	3,360	810	1,750	$<10^4$
01/02/90	112,000	3,600	1,040	1,690	$<10^4$
02/06/90	289,000	5,301	1,480	2,490	$<10^4$
05/11/93	24,000	2,461	390	1,020	0.86×10^5
05/18/94	23,000	1,430	300	770	ND[c]
09/11/95	23,000	2,312	160	1,500	1.21×10^5
09/13/95	12,000	532	10	230	1.84×10^5
09/27/95	20,000	4,508	180	4,060	1.65×10^5
10/19/95	12,000	3,854	120	2,960	1.21×10^5
11/16/95	5,000	5,616	170	4,660	1.05×10^5
12/19/95	20,000	1,920	60	1,520	1.19×10^5
01/23/96	12,000	4,366	90	3,150	1.07×10^5
02/13/96	12,000	11,232	110	6,510	0.90×10^5

[a] Lymphocyte subset counts in peripheral blood were determined by FACScan analysis as described previously (Ahmed-Ansari et al. [1989] Am. J. Primatol. 17:107-131).
[b] Virus levels were determined using the Chiron B-DNA assay as directed by the manufacturer. Plasma samples were stored at - 80°C until use.
[c] ND, not done (no plasma sample was available for this date).

Table 2A. Alignment of HIV-1_{JC} (SEQ ID NO:2) and HIV-1_{NDK} env (SEQ ID NO:3) Protein Sequences. The consensus sequence corresponds to SEQ ID NO:6.

```
HIVJCENV     1 MRvkEncqhlwrwgWKWGIMLLGMLMiCSAtEkLWVTVYYGVPvWKEtTTTLFCASDAKAY
HIVNDKENV    1 MRarE kerncqnlWKWGIMLLGMLMtCSAaEdLWVTVYYGVPiWKEaTTTLFCASDAKAY
consensus      MR--En--------WKWGIMLLGMLM-CSA-E-LWVTVYYGVP-WKE-TTTLFCASDAKAY

HIVJCENV    62 eeEvHNvWATHACVPTDPNPQEIvLaNVTEdFNMWKNeMVEQMHtDIISLWDeSLKPCVKL
HIVNDKENV   61 kkEaHNiWATHACVPTDPNPQEIeLeNVTEnFNMWKNnMVEQMHeDIISLWDqSLKPCVKL
consensus      --E-HN-WATHACVPTDPNPQEI-L-NVTE-FNMWKN-MVEQMH-DIISLWD-SLKPCVKL

HIVJCENV   123 TPLCVTLNCTDlknEtktNSsdaNsnsgEimgnEeiKNCSFNVstgapgkvqkeYaLfyal
HIVNDKENV  122 TPLCVTLNCTD   E lrNS kgNgkveE   eEkrKNCSFNVrdkreqvyalfYkL div
consensus      TPLCVTLNCTDlknEt--NSs--N----Eimg-E--KNCSFNV----------Y-Lf---

HIVJCENV   184 dIvsikNenNSTshmLtsCnTSvsTQACPKvSFEPIPIHyCAPAGFAILKCnDKKFNGTGP
HIVNDKENV  174 pI dnnNrtNSTnyrLinCdTStiTQACPKiSFEPIPIHfCAPAGFAILKCrDKKFNGTGP
consensus      -Iv---N--NST---L--C-TS--TQACPK-SFEPIPIH-CAPAGFAILKC-DKKFNGTGP

HIVJCENV   245 CnNVSTVQCTHGIRPVVSTQLLLNGSvAEEEvvlRSaNfsdNaKTIIVQLNhSveItCTRP
HIVNDKENV  234 CsNVSTVQCTHGIRPVVSTQLLLNGSlAEEEiiiRSeNltnNvKTIIVQLNaSivInCTRP
consensus      C-NVSTVQCTHGIRPVVSTQLLLNGS-AEEE---RS-N---N-KTIIVQLN-S--I-CTRP

HIVJCENV   306 nynetkkirIhrgygrsfvT  vrKlGdrkQAHCtmnRtkWdnALkQiAsKLreqfNKTaI
HIVNDKENV  295 ykytrqrtsIglrqslytiTgkkkKtGyigQAHCkisRaeWnkALqQvAtKLgnllNKTtI
consensus      ---------I---------Tgk--K-G---QAHC---R--W--AL-Q-A-KL----NKT-I

HIVJCENV   365 iFnrSSGGDlEIemHsfNCGGelFYCNTtkLFNSTWNeTtesngkgeniTLPCRIrQfVNm
HIVNDKENV  356 tFkpSSGGDpEItsHmlNCGGdfFYCNTsrLFNSTWNqTnstgfnngtvTLPCRIkQiVNl
consensus      -F--SSGGD-EI--H--NCGG--FYCNT--LFNSTWN-T----------TLPCRI-Q-VN-

HIVJCENV   426 WQkVGKAMYAPPsdGqIrCtSNITGLLLTRDGGhndNNtnnETfRPGrGDMRDNWRSELYK
HIVNDKENV  417 WQrVGKAMYAPPieGlIkCsSNITGLLLTRDGG  aNNsshETiRPGgGDMRDNWRSELYK
consensus      WQ-VGKAMYAPP--G-I-C-SNITGLLLTRDGGhn-NN---ET-RPG-GDMRDNWRSELYK
```

Table 2A. (Continued)

```
HIVJCENV  487 YKViKIEPlGVAPTKAkRRVVqREKRAvGmvGAmFLGFLGAAGSTMGAASlTLTVQARQLl
              ||| |||| ||||||| |||| ||||| |  || |||||||||||||||| |||||||||
HIVNDKENV 476 YKVvKIEPiGVAPTKArRRVVeREKRAiG lGAvFLGFLGAAGSTMGAASvTLTVQARQLm

consensus     YKV-KIEP-GVAPTKA-RRVV-REKRA-Gm-GA-FLGFLGAAGSTMGAAS-TLTVQARQL-


HIVJCENV  548 SGIVqQQNNLLRAIEAQQHLLQLTVWGIKQLQARVLAVERYLkDQQLLGIWGCSGklICTT
              |||| |||||||||||||||||||||||||||||||||||| |||||||||||  ||||
HIVNDKENV 536 SGIVhQQNNLLRAIEAQQHLLQLTVWGIKQLQARVLAVERYLrDQQLLGIWGCSGrhICTT

consensus     SGIV-QQNNLLRAIEAQQHLLQLTVWGIKQLQARVLAVERYL-DQQLLGIWGCSG--ICTT


HIVJCENV  609 tVPWNaSWSNkSLDqIWnNMTWmEWdREIaNYTnLIhhLIEESQnQQEKNEqELLELDKWA
               |||| |||| ||| || |||| || ||| ||| ||  |||||| |||||| |||||||||
HIVNDKENV 597 nVPWNsSWSNrSLDeIWqNMTWmEWeREIdNYTgLIysLIEESQiQQEKNEkELLELDKWA

consensus     -VPWN-SWSN-SLD-IW-NMTW-EW-REI-NYT-LI--LIEESQ-QQEKNE-ELLELDKWA


HIVJCENV  670 SLWsWFdIsnWLWYIKiFIMIVaGLvGLRIVFAVLSiVNRVRQGYSPLSFQThfPaPRGPD
              ||| || |  |||||| ||||| || |||||||||| |||||||||||||||  | |||||
HIVNDKENV 658 SLWnWFsItkWLWYIKlFIMIVgGLiGLRIVFAVLSvVNRVRQGYSPLSFQTllPvPRGPD

consensus     SLW-WF-I--WLWYIK-FIMIV-GL-GLRIVFAVLS-VNRVRQGYSPLSFQT--P-PRGPD


HIVJCENV  731 RPdgIEgEGGERdRDRSvRLVdGflALlWeDLRNLCLFSYHRLRDllLIvtRIVELLGRRG
              ||  || ||||| |||| ||| |  || | |||||||||||||||  ||  ||||||||||
HIVNDKENV 719 RPeeIEeEGGERgRDRSiRLVnGlfALfWdDLRNLCLFSYHRLRDsiLIaaRIVELLGRRG

consensus     RP--IE-EGGER-RDRS-RLV-G--AL-W-DLRNLCLFSYHRLRD--LI--RIVELLGRRG


HIVJCENV  792 WEALKYLWsLLQYWSQELkNSAvnLfnTtAIvVAEgTDRiIEVVQRlCRAILhiPRRIRQG
              |||||||| ||||||||| |||  |  | || ||| ||| |||||| |||||  |||||||
HIVNDKENV 780 WEALKYLWnLLQYWSQELrNSAssLldTiAIaVAErTDRvIEVVQRaCRAILnvPRRIRQG

consensus     WEALKYLW-LLQYWSQEL-NSA--L--T-AI-VAE-TDR-IEVVQR-CRAIL--PRRIRQG


HIVJCENV  853 LERiLL
              ||| ||
HIVNDKENV 841 LERlLL

consensus     LER-LL
```

Alignment score = 541.00

Table 2B. Alignment of HIV-1_{JC} and HIV-1_{LAI} env Protein Sequences (SEQ ID NO:2 and SEQ ID NO:4, respectively). The consensus sequence corresponds to SEQ ID NO:7.

```
HIVJCENV      1 MRVKEncQHLWRWGWKWGiMLLGmLMICSATEKLWVTVYYGVPVWKEtTTTLFCASDAKAY
                |||||  |||||||||||  ||||  |||||||||||||||||||||||  ||||||||||||
HIVlaienvt    1 MRVKEkyQHLWRWGWKWGtMLLGiLMICSATEKLWVTVYYGVPVWKEaTTTLFCASDAKAY
consensus       MRVKE--QHLWRWGWKWG-MLLG-LMICSATEKLWVTVYYGVPVWKE-TTTLFCASDAKAY

HIVJCENV     62 eeEVHNVWATHACVPTDPNPQEiVLaNVTEdFNMWKNeMVEQMHtDIISLWDeSLKPCVKL
                ||||||||||||||||||||  ||  ||||  ||||||  ||||||  |||||||  ||||||||
HIVlaienvt   62 dtEVHNVWATHACVPTDPNPQEvVLvNVTEnFNMWKNdMVEQMHeDIISLWDqSLKPCVKL
consensus       --EVHNVWATHACVPTDPNPQE-VL-NVTE-FNMWKN-MVEQMH-DIISLWD-SLKPCVKL

HIVJCENV    123 TPLCVtLnCTDLkNeTkTNSSdaNSnSGEiM gneEIKNCSFNvSTgapGKVQKEYalFYa
                |||||  |  ||||  |  |  ||||  ||  |||  |     ||||||||  ||   |||||||  ||
HIVlaienvt  123 TPLCVsLkCTDLgNaTnTNSSntNSsSGEmMmekgEIKNCSFNiSTsirGKVQKEYafFYk
consensus       TPLCV-L-CTDL-N-T-TNSS--NS-SGE-Mm---EIKNCSFN-ST---GKVQKEY--FY-

HIVJCENV    183 LDIvsIkneNnsTShmLTSCNTSVsTQACPKVSFEPIPIHYCAPAGFAILKCNdKkFNGTG
                |||  |   |  ||  ||||||||  |||||||||||||||||||||||||  |  |||||
HIVlaienvt  184 LDI  IpidNdtTSytLTSCNTSViTQACPKVSFEPIPIHYCAPAGFAILKCNnKtFNGTG
consensus       LDIvsI---N--TS--LTSCNTSV-TQACPKVSFEPIPIHYCAPAGFAILKCN-K-FNGTG

HIVJCENV    244 PCnNVSTVQCTHGIRPVVSTQLLLNGSvAEEEVVlRSANFsDNAKTIIVQLNhSVEItCTR
                ||  ||||||||||||||||||||||  ||||||  |||||  ||||||||||  ||||  |||
HIVlaienvt  243 PCtNVSTVQCTHGIRPVVSTQLLLNGSlAEEEVViRSANFtDNAKTIIVQLNqSVEInCTR
consensus       PC-NVSTVQCTHGIRPVVSTQLLLNGS-AEEEVV-RSANF-DNAKTIIVQLN-SVEI-CTR

HIVJCENV    305 PNyNetKkIRIhRGyGRsFVTvrKlGdrkQAHCtmnRtKWdnaLKQIASKLREQFnNktaI
                ||  |  | |||  ||  ||  |||  | |  ||||  | ||  ||||||||||||  |  |
HIVlaienvt  304 PNnNtrKsIRIqRGpGRaFVTigKiGnmrQAHCnisRaKWnatLKQIASKLREQFgNnktI
consensus       PN-N--K-IRI-RG-GR-FVT--K-G---QAHC---R-KW---LKQIASKLREQFgN---I

HIVJCENV    365 IFnrSSGGDlEIemHSFNCGGElFYCNtTkLFNSTW NeT  TE SN gkG enITLPCRI
                ||  |||||  ||  ||||||||  ||||  |  ||||||  |  |  ||  ||  |  |||||||
HIVlaienvt  365 IFkqSSGGDpEIvtHSFNCGGEfFYCNsTqLFNSTWfNsTwsTEgSNnteGsdtITLPCRI
consensus       IF--SSGGD-EI--HSFNCGGE-FYCN-T-LFNSTWfN-TwsTEgSNn--Gs--ITLPCRI

HIVJCENV    420 rQFvNMWQkVGKAMYAPPsdGQIRCtSNITGLLLTRDGGhNdNNtnnEtFRPGrGDMRDNW
                  ||  ||||  |||||||||  |||||  ||||||||||||  |  ||  |  ||||  |||||||
HIVlaienvt  426 kQFiNMWQeVGKAMYAPPisGQIRCsSNITGLLLTRDGG N NNngsEiFRPGgGDMRDNW
consensus       -QF-NMWQ-VGKAMYAPP--GQIRC-SNITGLLLTRDGGhNdNN---E-FRPG-GDMRDNW
```

Table 2B. (Continued)

```
HIVJCENV  481 RSELYKYKViKIEPLGVAPTKAKRRVVQREKRAVGmvGAmFLGFLGAAGSTMGAaSlTLTV
              ||||||||| ||||||||||||||||||||||||  || |||||||||||||| | ||||
HIVlaienvt  485 RSELYKYKVvKIEPLGVAPTKAKRRVVQREKRAVG iGAlFLGFLGAAGSTMGArSmTLTV

consensus       RSELYKYKV-KIEPLGVAPTKAKRRVVQREKRAVGm-GA-FLGFLGAAGSTMGA-S-TLTV


HIVJCENV  542 QARQLLSGIVQQQNNLLRAIEAQQHLLQLTVWGIKQLQARvLAVERYLKDQQLLGIWGCSG
              |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
HIVlaienvt  545 QARQLLSGIVQQQNNLLRAIEAQQHLLQLTVWGIKQLQARiLAVERYLKDQQLLGIWGCSG

consensus       QARQLLSGIVQQQNNLLRAIEAQQHLLQLTVWGIKQLQAR-LAVERYLKDQQLLGIWGCSG


HIVJCENV  603 KLICTTtVPWNASWSNKSLdQIWNNMTWmEWDREIaNYTnLIHhLIEESQNQQEKNEQELL
              |||||| |||||||||||| |||||||| |||||| ||| ||| ||||||||||||||||
HIVlaienvt  606 KLICTTaVPWNASWSNKSLeQIWNNMTWmEWDREInNYTsLIHsLIEESQNQQEKNEQELL

consensus       KLICTT-VPWNASWSNKSL-QIWNNMTW-EWDREI-NYT-LIH-LIEESQNQQEKNEQELL


HIVJCENV  664 ELDKWASLWsWFdIsNWLWYIKIFIMIVaGLVGLRIVFAVLSIVNRVRQGYSPLSFQTHfP
              ||||||||| || | |||||||||||| |||||||||||||||||||||||||||||| |
HIVlaienvt  667 ELDKWASLWnWFnItNWLWYIKIFIMIVgGLVGLRIVFAVLSIVNRVRQGYSPLSFQTHlP

consensus       ELDKWASLW-WF-I-NWLWYIKIFIMIV-GLVGLRIVFAVLSIVNRVRQGYSPLSFQTH-P


HIVJCENV  725 aPRGPDRPdGIEgEGGERDRDRSvRLVdGfLALlWeDLRnLCLFSYHRLRDLLLIVTRIVE
              ||||||| ||| |||||||||| ||| | ||| | ||| ||||||||||||||||||||
HIVlaienvt  728 tPRGPDRPeGIEeEGGERDRDRSiRLVnGsLALiWdDLRsLCLFSYHRLRDLLLIVTRIVE

consensus       -PRGPDRP-GIE-EGGERDRDRS-RLV-G-LAL-W-DLR-LCLFSYHRLRDLLLIVTRIVE


HIVJCENV  786 LLGRRGWEALKYlWsLLQYWSQELKNSAVnLfNtTAIvVAEGTDRiIEVVQrlCRAIlHIP
              |||||||||||| | |||||||||||||| | | ||| ||||||| |||||  |||| |||
HIVlaienvt  789 LLGRRGWEALKYwWnLLQYWSQELKNSAVsLlNaTAIaVAEGTDRvIEVVQgaCRAIrHIP

consensus       LLGRRGWEALKY-W-LLQYWSQELKNSAV-L-N-TAI-VAEGTDR-IEVVQ--CRAI-HIP


HIVJCENV  847 RRIRQGLERiLL
              |||||||||  ||
HIVlaienvt  850 RRIRQGLERiLL

consensus       RRIRQGLER-LL


Alignment score = 657.00
```

Table 2C. Alignment of HIV-1_{JC} and HIV-1_{SF2} env Protein Sequences (SEQ ID NO:2 and SEQ ID NO:5, respectively). The consensus sequence corresponds to SEQ ID NO:8.

```
HIVJCENV     1 MrVKencqhlwrwgWkWGimLLGMLMICSATEKLWVTVYYGVPVWKEtTTTLFCASDAkAY
HIVsf2envt   1 MkVK gtrrnyqhlWrWGtlLLGMLMICSATEKLWVTVYYGVPVWKEaTTTLFCASDArAY
consensus      M-VKe---------W-WG--LLGMLMICSATEKLWVTVYYGVPVWKE-TTTLFCASDA-AY

HIVJCENV    62 eeEVHNVWATHACVPTDPNPQEiVLaNVTEdFNMWKNeMVEQMhtDIISLWDeSLKPCVKL
HIVsf2envt  61 dtEVHNVWATHACVPTDPNPQEvVLgNVTEnFNMWKNnMVEQMqeDIISLWDqSLKPCVKL
consensus      --EVHNVWATHACVPTDPNPQE-VL-NVTE-FNMWKN-MVEQM--DIISLWD-SLKPCVKL

HIVJCENV   123 TPLCVTLNCTDLkneTkTNSSdaNsnsgeimgneeiknсsfnvstgapgKvqkeyalfyal
HIVsf2envt 122 TPLCVTLNCTDLgkaTnTNSS  NwkeeikgeiknсsfnittsirdkiqKenalfrnldvv
consensus      TPLCVTLNCTDL---T-TNSSdaN-----------------------K----------

HIVJCENV   184 dIvsiknenNsTshmLtsCNtSVsTQACPKVSFEPIPIHYCaPAGFAILKCNdKkFNGtGP
HIVsf2envt 181 pIdnastttNyTnyrLihCNrSViTQACPKVSFEPIPIHYCtPAGFAILKCNnKtFNGkGP
consensus      -I-------N-T---L--CN-SV-TQACPKVSFEPIPIHYC-PAGFAILKCN-K-FNG-GP

HIVJCENV   245 CnNVSTVQCTHGIRPvVSTQLLLNGSvAEEEVVlRSaNFsdNAKTIIVQLNhSVeItCTRP
HIVsf2envt 242 CtNVSTVQCTHGIRPiVSTQLLLNGSlAEEEVViRSdNFtnNAKTIIVQLNeSVaInCTRP
consensus      C-NVSTVQCTHGIRP-VSTQLLLNGS-AEEEVV-RS-NF--NAKTIIVQLN-SV-I-CTRP

HIVJCENV   306 NyNetKkIrIhrGygrsfvTvRklGDrkqAHCtmnRtkWdNaLkQIasKLREQFnNktaIi
HIVsf2envt 303 NnNtrKsIyIgpG rafhtTgRiiGDirkAHCnisRaqWnNtLeQIvkKLREQFgNnktIv
consensus      N-N--K-I-I--Gy-----T-R--GD---AHC---R--W-N-L-QI--KLREQFgN---I-

HIVJCENV   366 FNrSSGGDlEIeMHSFNCgGElFYCNTTkLFNsTW netTEsngkgenItLPCRIrQfvNM
HIVsf2envt 363 FNqSSGGDpEIvMHSFNCrGEfFYCNTTqLFNnTWrlnhTEgtkgndtIiLPCRIkQiiNM
consensus      FN-SSGGD-EI-MHSFNC-GE-FYCNTT-LFN-TWr---TE-------I-LPCRI-Q--NM

HIVJCENV   426 WQkVGKAMYAPPsdGQIrCtSNITGLLLTRDGGhNdnNtnnEtFRPGrGDMRDNWRSELYK
HIVsf2envt 424 WQeVGKAMYAPPigGQIsCsSNITGLLLTRDGGtNvtN dtEvFRPGgGDMRDNWRSELYK
consensus      WQ-VGKAMYAPP--GQI-C-SNITGLLLTRDGG-N--Nt--E-FRPG-GDMRDNWRSELYK
```

Table 2C. (Continued)

```
HIVJCENV   487 YKVIKIEPLGvAPTKAKRRVVQREKRAVGmVGAMFLGFLGAAGSTMGAaSLTLTVQARQLL
               |||||||||| |||||||||||||||||| |||||||||||||||||| ||||||||||||
HIVsf2envt 484 YKVIKIEPLGiAPTKAKRRVVQREKRAVGiVGAMFLGFLGAAGSTMGAvSLTLTVQARQLL

consensus      YKVIKIEPLG-APTKAKRRVVQREKRAVG-VGAMFLGFLGAAGSTMGA-SLTLTVQARQLL


HIVJCENV   548 SGIVQQQNNLLRAIEAQQHLLQLTVWGIKQLQARVLAVERYLkDQQLLGIWGCSGKLICTT
               |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
HIVsf2envt 545 SGIVQQQNNLLRAIEAQQHLLQLTVWGIKQLQARVLAVERYLrDQQLLGIWGCSGKLICTT

consensus      SGIVQQQNNLLRAIEAQQHLLQLTVWGIKQLQARVLAVERYL-DQQLLGIWGCSGKLICTT


HIVJCENV   609 tVPWNASWSNKSLdqIWnNMTWmeWdREIaNYTNlIhhLiEESQNQQEKNEQELLELDKWA
               |||||||||||  || ||||  | ||| |||| |  | ||||||||||||||||||||
HIVsf2envt 606 aVPWNASWSNKSLedIWdNMTWmqWeREIdNYTNtIytLlEESQNQQEKNEQELLELDKWA

consensus      -VPWNASWSNKSL--IW-NMTW--W-REI-NYTN-I--L-EESQNQQEKNEQELLELDKWA


HIVJCENV   670 SLWsWFdIsNWLWYIKIFIMIVaGLVGLRIVFAVLSIVNRVRQGYSPLSFQThfPaPRGPD
               ||| || | |||||||||||| ||||||||||||||||||||||||||  | |||||
HIVsf2envt 667 SLWnWFsItNWLWYIKIFIMIVgGLVGLRIVFAVLSIVNRVRQGYSPLSFQTrlPvPRGPD

consensus      SLW-WF-I-NWLWYIKIFIMIV-GLVGLRIVFAVLSIVNRVRQGYSPLSFQT--P-PRGPD


HIVJCENV   731 RPDGIEgEGGERDRDRSVRLVDGFLALlWEDLRnLCLFSYhRLRDLLLIvtRiVElLGrRG
               |||||| |||||||||||||||||||| ||||| |||||| ||||||||  | || || ||
HIVsf2envt 728 RPDGIEeEGGERDRDRSVRLVDGFLALiWEDLRsLCLFSYrRLRDLLLIaaRtVEiLGhRG

consensus      RPDGIE-EGGERDRDRSVRLVDGFLAL-WEDLR-LCLFSY-RLRDLLLI--R-VE-LG-RG


HIVJCENV   792 WEALKYlWSLLQYWsQELKNSAVnlfNtTAIvVaEGTDRiIEVvQRlcRAILHIpRRIRQG
               |||||| ||||||| ||||||||   | ||| | ||||| ||| ||  |||||| ||||||
HIVsf2envt 789 WEALKYwWSLLQYWiQELKNSAVswlNaTAIaVtEGTDRvIEVaQRayRAILHIhRRIRQG

consensus      WEALKY-WSLLQYW-QELKNSAV---N-TAI-V-EGTDR-IEV-QR--RAILHI-RRIRQG


HIVJCENV   853 LERiLL
               ||| ||
HIVsf2envt 850 LERlLL

consensus      LER-LL

Alignment score = 614.00
```

Table 3. HIV-1 Antibody Response of C534 After Transfusion From C455

Days Post Infection	Titer
0	0
7	0
14	400
21	1600
28	1600

Table 4. Comparative Analyses of Chimpanzee HIV-1 Isolates With Inoculating Viruses

		HIV-1_{JC16} vs:				Mutations unique to JC/NC[2]
		LAI	SF2	NDK	NC7	
Pol	% homology[1]	96.6	94.8	94.0	98.8	-
	insertions	12	0	0	0	0
	deletions	0	0	1	0	0
	point mutations	30	51	56	11	23
Gag	% homology[1]	94.2	92.2	88.8	97.4	-
	insertions	12	5	1	2	0
	deletions	0	3	81	0	2
	point mutations	25	33	42	9	20
Vif	% homology[1]	92.2	86.0	86.5	96.9	-
	insertions	0	0	0	0	0
	deletions	0	0	0	0	0
	point mutations	14	26	25	5	10
Vpr	% homology[1]	88.7	96.9	89.7	96.6	-
	insertions	0	0	0	0	0
	deletions	1	0	1	0	0
	point mutations	7	2	6	2	0
Tat	% homology[1]	88.4	82.5	74.4	97.0	-
	insertions	0	1	0	0	0
	deletions	15	0	15	0	0
	point mutations	10	18	22	3	4
Nef	% homology[1]	87.4	89.4	78.7	99.0	-
	insertions	0	4	1	0	0
	deletions	0	0	0	0	0
	point mutations	25	18	40	1	8
Rev	% homology[1]	84.5	82.9	75.2	96.6	-
	insertions	1	1	1	0	0
	deletions	1	0	0	0	0
	point mutations	18	19	28	4	11
Env	% homology[1]	81.9	76.1	71.4	97.6	-
	insertions	9	9	11	0	0
	deletions	7	13	24	1	2
	point mutations	138	173	203	17	78
Vpu	% homology[1]	79.3	63.9	67.1	98.8	-
	insertions	0	0	0	0	0
	deletions	1	1	1	0	0
	point mutations	13	24	24	0	10
LTR	% homology[1]	92.1	93.9	89.1	98.7	-
	insertions	1	3	1	0	0
	deletions	1	3	1	0	1
	point mutations	48	29	52	10	14

[1]Amino acid homology
[2]Amino acids or nucleotides not observed in parental viruses
[3]Nucleotide identity

Table 5

```
                                |---------MBD--------|-> matrix protein (p17)
HIV-1JC16 gag MGARASVLSGGDLDRWEKIRLRPGGKKKYMLKHIVWASRELERFAVNPGLLETSEGCRQVLEQLQPSLQT 70
HIV-1NC7 gag  ..........................................................A.....I.......... 70
HIV-1LAI gag  ..........E.................K.....................................I.G..... 70
HIV-1SF2 gag  ..........E..K.............K.....................................I.G..... 70
HIV-1NDK gag  ..........K..T..R.........A...LI.........TL..........K.IIG.....I.. 70
                                                                                |->
HIV-1JC16 gag GSEELRSLYNTIATLYGVHQKIEVKDTKEALDKIEEEQNKSKKKAQQAAA-------DTGHSSQVSQNYP 133
HIV-1NC7 gag  ................C.............E...........V.....AT-----G..N.......... 135
HIV-1LAI gag  ..........V....C...R..I.....................------------------------- 133
HIV-1SF2 gag  ..........V....C...R.D.......E.....................AAGTGN------------ 135
HIV-1NDK gag  ....I.....V....C..ER.........VE.M.........T.....------------------- 130
              capsid protein (p24)
HIV-1JC16 gag IVQNLQGQMVHQAISPRTLNAWVKVIEEKAFSPEVIPMFSALSEGATPQDLNTMLNTVGGHQAAMQMLKE 203
HIV-1NC7 gag  ........................V..................................................... 205
HIV-1LAI gag  ....I...................V..................................................... 203
HIV-1SF2 gag  ........................V..................................................... 205
HIV-1NDK gag  ........................V..................................................... 200

HIV-1JC16 gag TINEEAAEWDRVHPVQAGLIAPGQMREPRGSDIAGTTSTLQEQIGWMTSNPPIPVGEIYKRWIILGLNKI 273
HIV-1NC7 gag  ..........................................................N..................... 275
HIV-1LAI gag  ................H..P.....................................N..................... 273
HIV-1SF2 gag  ................H..P.....................................N..................... 275
HIV-1NDK gag  ...D.......L...H..PV.....................A..................................... 270

HIV-1JC16 gag VRMYSPISILDIRQGPKEPFRDYVDRFYKTLRAEQASQEVKNWMTETLLVQNANPDCKTILKALGPGATL 343
HIV-1NC7 gag  ..........................................................A...... 345
HIV-1LAI gag  ......T.....................................D.....................A...... 343
HIV-1SF2 gag  ......T.....................................D.....................Q...... 345
HIV-1NDK gag  ......V.....................................D..................... 340
                                p24 <-|     |------p2------|-> Nucleocapsid  [  zinc finger  ]  (p7)
HIV-1JC16 gag EEMMTACQGVGGPGHKARVLAEAMSQVT-NSAAIMMQRGNFKNQRKMVKCFNCGKEGHVARNCRAPRKKG 412
HIV-1NC7 gag  ..............................-.....T.......R....I...........I.......... 414
HIV-1LAI gag  ..............................-.P.N.......R....T...........I.K........ 412
HIV-1SF2 gag  ..............................GSAT.V.......GP..SI...........T.K........ 414
HIV-1NDK gag  ..............................                                          410
```

Table 5 (continued)

```
                 [   zinc finger   ] p7  <-|--------p1---------|-> p6
HIV-1JC16 gag CWKCGKEGHQMKDCIESQANFLGKIWPSHKGRPGNFLQSRPEPTAPP----------EESLRSGIETT 470
HIV-1NC7 gag  ...............R.....................................................  472
HIV-1LAI gag  ...............R.....................FLQSRPEPTAPP...F...V...  482
HIV-1SF2 gag  ...............R.....Y.....................................  472
HIV-1NDK gag  ..R..R.....T.R.....Y.....................F.F.E.K.  420
              .....R....
                                                                  p6 <-|
HIV-1JC16 gag TPSQKQEPIDKEVYPLTSLRSLFGNDPSSQ                              501
HIV-1NC7 gag  ..............................                               503
HIV-1LAI gag  ...........L..................                               513
HIV-1SF2 gag  ...........L..................                               503
HIV-1NDK gag                                                               420
```

Table 5 (continued)

```
                |--MTD---|
HIV-1JC16 nef MGGKWSKSSIVGWPTIRERMKRAG----PAADGVGAASRDLEKHGAITSSNTAATNADCA 56
HIV-1NC7 nef  ..........................................R.................. 56
HIV-1LAI nef  .........V.....V....R..E-------------------------------A.. 56
HIV-1SF2 nef  .......R.MG..SA.....R..EPRAE........V.....................S..DT.. 60
HIV-1NDK nef  .........L....A....IRKTD-----------V.....................S..DT.. 56
                              LCK binding domain
                              |--------------|
HIV-1JC16 nef WLEAQEE-EEVGFPVRPQVPLRPMTYKAGIDLSHFLKEKGGLEGLVWSQRRQDILDLWIY 115
HIV-1NC7 nef  .......-.....................................................  115
HIV-1LAI nef  ..............T.............AV..............IH.............  115
HIV-1SF2 nef  .......-.....................AL.I...........I......E.......  119
HIV-1NDK nef  .......S.....................EAV..............I..KK..E.....V.  116

HIV-1JC16 nef HTQGYFPDWQNYTPGPGIRYPLTFGWCFKLVPVEPDKVEEANEGEDNILLHPMCLHGMED 175
HIV-1NC7 nef  .....................................................................  175
HIV-1LAI nef  ..................V.........Y..............K..NTS....VS....D.  175
HIV-1SF2 nef  ..................................E.........N.S.....S......  179
HIV-1NDK nef  N...I.....................Q....D.QE....T.R...C......QQ....  176

HIV-1JC16 nef AEKEVLVWRFDSKLAFHHVARELHPEYYKDC                              207
HIV-1NC7 nef  ..............................                              207
HIV-1LAI nef  P.R...E.....R..............F.N.                              207
HIV-1SF2 nef  ..................M.............                              211
HIV-1NDK nef  P.RQ..M...N.R..LE.K.......F....                              208
```

Table 5 (continued)

```
                                                signal peptide <-|->gp120 (SU)
HIV-1JC16 env MRVKE---NCQHLWRWGWKWGIMLLGMLMICSATEKLWVTVYYGVPVWKETTTTLFCASD 57
HIV-1NC7 env  .........Y.....................................A..........  57
HIV-1LAI env  .........KY..........T....I.....................A..........  56
HIV-1SF2 env  .K..GTRR.Y..........TL.........................A..........  56
HIV-1NDK env  ..AR.KER...N.----------------T...A.D.........I...A..........  56

HIV-1JC16 env AKAYEEEVHNVWATHACVPTDPNPQEIVLANVTEDFNMWKNEMVEQMHTDIISLWDESLK 117
HIV-1NC7 env  ..............................................................  117
HIV-1LAI env  ....DT.....................V..V....N......D......E.......Q...  117
HIV-1SF2 env  ....DT.....................V..G....N......N.....QE.......Q...  116
HIV-1NDK env  .R..DT.....................V..G....N......N.....E.......Q...  116
HIV-1NDK env  ....KK.A..I.................E.E....N.....|----------------V2--
                        |---------------V1-------------------|
HIV-1JC16 env PCVKLTPLCVTLNCTDLKNETKTNSSDANSNSGEIM-GNEEIKNCSFNVSTGAPGKVQKE 176
HIV-1NC7 env  ..............................................................  176
HIV-1LAI env  ..........S.K....G.A.N....NT..S...M.MEKG........I..SIR......  177
HIV-1SF2 env  ..................GKA.N....NWKEEI------KG........IT.SIRD.I...  170
HIV-1NDK env  ..................ELRNS.G.GK---------VEEE.KR.............RD.REQV  163
              --------------V2-------------------|
HIV-1JC16 env YALFYALDIVSIKNENNSTSH---MLTSCNTSVSTQACPKVSFEPIPIHYCAPAGFAILK 233
HIV-1NC7 env  ..............................................................  233
HIV-1LAI env  ..F..K...IP.D.DTT.-----YT.......I.............................  232
HIV-1SF2 env  N...RN..V.P.D.ASTT.NYTNYR.IH..R..I.....................T.......  230
HIV-1NDK env  .....K....P.D.N.RTN.-TNYR.IN.D..TI......I.........F..........  222

HIV-1JC16 env CNDKKFNGTGPCNNVSTVQCTHGIRPVVSTQLLLNGSVAEEEVVLRSANFSDNAKTIIVQ 293
HIV-1NC7 env  ..............................................L...............  293
HIV-1LAI env  ..N.T.......T.....................................I.....T.....  292
HIV-1SF2 env  ..N.T...K...T.....................I.....L......I..D..TN......  290
HIV-1NDK env  .R..........S.....................L....III..E.LTN.V......  282
                        |-----------------------V3------------------|
HIV-1JC16 env LNHSVEITCTRPNYNETKKIRIHRGYGRSFVTV----RKLGDRKQAHCTMNRTKWDNALK 349
HIV-1NC7 env  ..............R.....R.....R.....R.....R.....R.....R.....R..  349
HIV-1LAI env  ..Q....N.....N.TR.S...Q..P..A...IG---KI-.NMR....NIS.A..NAT..  348
HIV-1SF2 env  ..E..A.N.....N.TR.S.Y.--.P..A.H.TG----.II..IRK...NIS.AQ.N.T.E  345
HIV-1NDK env  ..A.IV.N....YKYTRQRTS.--.LRQ.LY.ITGKKK.T.YIG....KIS.AE.NK..Q  340
```

SUBSTITUTE SHEET (RULE 26)

Table 5 (continued)

```
                                                                                                  |--------------V4-------
HIV-1JC16 env QIASKLREQFNNKTAIIFNRSSGGDLEIEMHSFNCGGELFYCNTTKLFNSTW------NE 403
HIV-1NC7 env  ..........-.....................................................--------. 402
HIV-1LAI env  ..........G.NKT...KQ.....P..VT........F....S.Q......FNSTWST. 408
HIV-1SF2 env  ..........G.NKT...KQ.....P..VT.....R..F.....Q...N.------WRLN 400
HIV-1NDK env  ..VK......G.NKT.V..Q.....P..V......R..F.....Q...N.------WRLN 393
              .V.T...-GNLL...T.T.KP.....P..TS.ML....DF.....SR............Q
                                                  CD4 binding
                                                  |----------|                |---V5--
              --------V4---------|                                                        
HIV-1JC16 env TTESNGKGENITLPCRIRQFVNMWQKVGKAMYAPPSDGQIRCTSNITGLLLTRDGGHNDN 463
HIV-1NC7 env  ..........................................................GPSD 462
HIV-1LAI env  GSNNTEGSDT.......K..I....E.........IS.....S.............N.--. 467
HIV-1SF2 env  H..GTKGNDT.I.....K.II....E.........IG...S.S.............T.VT 460
HIV-1NDK env  .NSTGFNNGTV......K.I..L..R.........IE.L.K.S.............A.-- 451
                                                  gp120 (SU) <-|-> gp41 (TM)
              ---V5---|
HIV-1JC16 env NTNNETFRPGRGDMRDNWRSELYKYKVIKIEPLGVAPTKAKRRVVQREKRAVGMVGAMFL 523
HIV-1NC7 env  .KTDK.....G.....................................I-..L.. 522
HIV-1LAI env  .-GS.I....G.....................V.....................I-..L.. 525
HIV-1SF2 env  .-GS.I....G.....................I.....................I..... 519
HIV-1NDK env  .-DT.V....G.....................V....I.....R....E.....I.L-..V.. 510
              .SSH..I...G.....................V....I.....R....E.....I.L-..V..

HIV-1JC16 env GFLGAAGSTMGAASLTLTVQARQLLSGIVQQQNNLLRAIEAQQHLLQLTVWGIKQLQARV 583
HIV-1NC7 env  ..........................................................582
HIV-1LAI env  ..........................................................I 585
HIV-1SF2 env  ..........R.M.....................................................579
HIV-1NDK env  ..........V.....................................................570
              ..........V.....M....H.....

HIV-1JC16 env LAVERYLKDQQLLGIWGCSGKLICTTTVPWNASWSNKSLDQIWNNMTWMEWDREIANYTN 643
HIV-1NC7 env  ..........................................L.....................642
HIV-1LAI env  ..........................A.....E.....................N...S 645
HIV-1SF2 env  .....R....................A.....ED..D.....Q.E...D.... 639
HIV-1NDK env  .....R....................RH....N....S....R...E..Q.....E...D...G 630

HIV-1JC16 env LIHHLIEESQNQQEKNEQELLELDKWASLWSWFDISNWLWYIKIFIMIVAGLVGLRIVFA 703
HIV-1NC7 env  ..........................................................702
HIV-1LAI env  ...S......................N..N.T.....G.....................705
HIV-1SF2 env  T.YT.L....................N..S.T.....G.....................699
HIV-1NDK env  ..YS.....I.....K.........N..S.TK.....L.....G..I.....690
```

Table 5 (continued)

```
HIV-1JC16 env VLSIVNRVRQGYSPLSFQTHFPAPRGPDRPDGIEGEGGERDRDRSVRLVDGFLALLWEDL 763
HIV-1NC7 env  ............................................................ 762
HIV-1LAI env  ..................L.T.....E..E.........I...N.S...I.D.. 765
HIV-1SF2 env  ..................RL.V..........E.....................I.... 759
HIV-1NDK env  ...V..............LL.V.....EE..E.....G....I...N.LF..F.D.. 750

HIV-1JC16 env RNLCLFSYHRLRDLLLIVTRIVELLGRRGWEALKYLWSLLQYWSQELKNSAVNLFNTTAI 823
HIV-1NC7 env  ..................................W.....................R.. 822
HIV-1LAI env  ..................................W.N.............S.L.A... 825
HIV-1SF2 env  .S................................W.N.............S.L.A... 819
HIV-1NDK env  .S.....R.....AA.T..I..H.....W.....I.....SWL.A... 819
HIV-1NDK env  ..........SI..AA..................N.........R...SS.LD.I.. 810

HIV-1JC16 env VVAEGTDRIIEVVQRLCRAILHIPRRIRQGLERILL 860
HIV-1NC7 env  ..................................F.. 859
HIV-1LAI env  A.....V.....GA....R.............. 862
HIV-1SF2 env  A.T.....V...A..AY.....H.........L.. 856
HIV-1NDK env  A...R...V.....A.....NV.........L.. 847
```

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: EMORY UNIVERSITY

(ii) TITLE OF INVENTION: Human Immunodeficiency Viruses Causing AIDS in a Nonhuman Primate

(iii) NUMBER OF SEQUENCES: 23

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
(B) STREET: 5370 Manhattan Circle, Suite 201
(C) CITY: Boulder
(D) STATE: Colorado
(E) COUNTRY: US
(F) ZIP: 80303

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: WO
(B) FILING DATE: 23-JUN-1998
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 60/050,548
(B) FILING DATE: 23-JUN-1997

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 60/057,606
(B) FILING DATE: 04-SEP-1997

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Ferber M., Donna
(B) REGISTRATION NUMBER: 33,878
(C) REFERENCE/DOCKET NUMBER: 66-97 WO

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: (303) 499-8080
(B) TELEFAX: (303) 499-8089

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2577 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..2577

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
ATG AGA GTG AAG GAG AAC TGT CAG CAC TTG TGG AGA TGG GGG TGG AAA      48
Met Arg Val Lys Glu Asn Cys Gln His Leu Trp Arg Trp Gly Trp Lys
 1               5                  10                  15

TGG GGC ATC ATG CTC CTT GGG ATG TTA ATG ATC TGT AGT GCT ACA GAA      96
Trp Gly Ile Met Leu Leu Gly Met Leu Met Ile Cys Ser Ala Thr Glu
            20                  25                  30

AAA TTG TGG GTC ACA GTC TAT TAT GGG GTA CCT GTG TGG AAG GAA ACA     144
Lys Leu Trp Val Thr Val Tyr Tyr Gly Val Pro Val Trp Lys Glu Thr
        35                  40                  45

ACT ACC ACT CTA TTT TGT GCA TCA GAT GCT AAA GCA TAT GAA GAA GAG     192
Thr Thr Thr Leu Phe Cys Ala Ser Asp Ala Lys Ala Tyr Glu Glu Glu
    50                  55                  60

GTA CAT AAT GTT TGG GCC ACA CAT GCC TGT GTA CCC ACA GAC CCC AAC     240
Val His Asn Val Trp Ala Thr His Ala Cys Val Pro Thr Asp Pro Asn
65                  70                  75                  80

CCA CAA GAA ATA GTA TTG GCA AAT GTG ACA GAA GAT TTT AAC ATG TGG     288
Pro Gln Glu Ile Val Leu Ala Asn Val Thr Glu Asp Phe Asn Met Trp
                85                  90                  95

AAA AAT GAA ATG GTA GAA CAG ATG CAT ACT GAT ATA ATC AGT TTA TGG     336
Lys Asn Glu Met Val Glu Gln Met His Thr Asp Ile Ile Ser Leu Trp
            100                 105                 110

GAT GAA AGC CTA AAA CCA TGT GTA AAA TTA ACC CCA CTC TGT GTT ACT     384
Asp Glu Ser Leu Lys Pro Cys Val Lys Leu Thr Pro Leu Cys Val Thr
        115                 120                 125

TTA AAT TGC ACT GAT TTG AAG AAT GAA ACT AAG ACC AAT AGT AGT GAT     432
Leu Asn Cys Thr Asp Leu Lys Asn Glu Thr Lys Thr Asn Ser Ser Asp
    130                 135                 140

GCC AAT AGT AAT AGC GGG GAA ATA ATG GGG AAC GAA GAG ATA AAA AAT     480
Ala Asn Ser Asn Ser Gly Glu Ile Met Gly Asn Glu Glu Ile Lys Asn
145                 150                 155                 160

TGC TCT TTC AAT GTC AGC ACA GGC GCA CCA GGT AAG GTG CAG AAA GAA     528
Cys Ser Phe Asn Val Ser Thr Gly Ala Pro Gly Lys Val Gln Lys Glu
                165                 170                 175
```

```
TAT TCA CTT TTT TAT GCA CTT GAT ATA GTA TCA ATA AAG AAT GAA AAT          576
Tyr Ser Leu Phe Tyr Ala Leu Asp Ile Val Ser Ile Lys Asn Glu Asn
            180                 185                 190

AAT AGT ACC AGC CAT ATG TTG ACA AGT TGT AAC ACC TCA GTC AGT ACA          624
Asn Ser Thr Ser His Met Leu Thr Ser Cys Asn Thr Ser Val Ser Thr
        195                 200                 205

CAG GCC TGT CCA AAG GTA TCC TTT GAG CCA ATT CCC ATA CAT TAT TGT          672
Gln Ala Cys Pro Lys Val Ser Phe Glu Pro Ile Pro Ile His Tyr Cys
    210                 215                 220

GCC CCG GCT GGT TTT GCA ATT CTA AAA TGT AAT GAT AAG AAG TTC AAT          720
Ala Pro Ala Gly Phe Ala Ile Leu Lys Cys Asn Asp Lys Lys Phe Asn
225                 230                 235                 240

GGA ACA GGA CCA TGT AAC AAT GTC AGC ACA GTA CAA TGT ACA CAT GGA          768
Gly Thr Gly Pro Cys Asn Asn Val Ser Thr Val Gln Cys Thr His Gly
                245                 250                 255

ATT AGA CCA GTA GTG TCA ACT CAA CTG CTG TTA AAT GGC AGT GTA GCA          816
Ile Arg Pro Val Val Ser Thr Gln Leu Leu Leu Asn Gly Ser Val Ala
            260                 265                 270

GAA GAA GAG GTA GTA CTT AGA TCT GCC AAT TTC TCA GAC AAT GCT AAA          864
Glu Glu Glu Val Val Leu Arg Ser Ala Asn Phe Ser Asp Asn Ala Lys
        275                 280                 285

ACC ATA ATA GTA CAG CTG AAC CAC TCT GTA GAA ATT ACT TGT ACA AGA          912
Thr Ile Ile Val Gln Leu Asn His Ser Val Glu Ile Thr Cys Thr Arg
    290                 295                 300

CCC AAC TAC AAT GAA ACA AAG AAA ATC CGT ATC CAC AGA GGA TAT GGA          960
Pro Asn Tyr Asn Glu Thr Lys Lys Ile Arg Ile His Arg Gly Tyr Gly
305                 310                 315                 320

AGA TCA TTT GTT ACA GTA AGA AAA TTG GGA GAT AGG AAA CAA GCA CAT         1008
Arg Ser Phe Val Thr Val Arg Lys Leu Gly Asp Arg Lys Gln Ala His
                325                 330                 335

TGT ACC ATG AAT AGA ACG AAA TGG GAC AAC GCT TTA AAA CAG ATA GCT         1056
Cys Thr Met Asn Arg Thr Lys Trp Asp Asn Ala Leu Lys Gln Ile Ala
            340                 345                 350

AGC AAA TTA AGA GAA CAA TTT AAT AAA ACA GCA ATA ATC TTT AAC CGG         1104
Ser Lys Leu Arg Glu Gln Phe Asn Lys Thr Ala Ile Ile Phe Asn Arg
        355                 360                 365

TCC TCA GGA GGG GAC CTA GAA ATT GAA ATG CAC AGT TTT AAT TGC GGA         1152
Ser Ser Gly Gly Asp Leu Glu Ile Glu Met His Ser Phe Asn Cys Gly
    370                 375                 380

GGG GAA TTG TTC TAC TGT AAT ACA ACA AAA CTG TTT AAT AGT ACT TGG         1200
Gly Glu Leu Phe Tyr Cys Asn Thr Thr Lys Leu Phe Asn Ser Thr Trp
385                 390                 395                 400
```

```
AAT GAG ACT ACA GAG TCA AAT GGC AAG GGA GAA AAT ATC ACA CTC CCA      1248
Asn Glu Thr Thr Glu Ser Asn Gly Lys Gly Glu Asn Ile Thr Leu Pro
                405                 410                 415

TGC AGA ATA AGA CAA TTT GTA AAC ATG TGG CAG AAA GTA GGA AAA GCA      1296
Cys Arg Ile Arg Gln Phe Val Asn Met Trp Gln Lys Val Gly Lys Ala
            420                 425                 430

ATG TAT GCC CCT CCC AGC GAT GGA CAA ATT AGG TGT ACA TCA AAT ATT      1344
Met Tyr Ala Pro Pro Ser Asp Gly Gln Ile Arg Cys Thr Ser Asn Ile
        435                 440                 445

ACT GGG CTA CTA TTA ACA AGA GAT GGG GGT CAT AAT GAT AAC AAC ACT      1392
Thr Gly Leu Leu Leu Thr Arg Asp Gly Gly His Asn Asp Asn Asn Thr
    450                 455                 460

AAC AAC GAG ACC TTC AGA CCG GGA AGA GGA GAT ATG AGG GAC AAT TGG      1440
Asn Asn Glu Thr Phe Arg Pro Gly Arg Gly Asp Met Arg Asp Asn Trp
465                 470                 475                 480

AGA AGT GAA TTA TAT AAA TAT AAA GTA ATA AAA ATT GAA CCA TTA GGA      1488
Arg Ser Glu Leu Tyr Lys Tyr Lys Val Ile Lys Ile Glu Pro Leu Gly
                485                 490                 495

GTA GCA CCC ACC AAG GCA AAG AGA AGA GTG GTG CAG AGA GAA AAA AGA      1536
Val Ala Pro Thr Lys Ala Lys Arg Arg Val Val Gln Arg Glu Lys Arg
            500                 505                 510

GCA GTG GGA ATG GTA GGA GCT ATG TTC CTT GGG TTC TTG GGA GCA GCA      1584
Ala Val Gly Met Val Gly Ala Met Phe Leu Gly Phe Leu Gly Ala Ala
        515                 520                 525

GGA AGC ACT ATG GGC GCA GCG TCA TTG ACG CTG ACG GTA CAG GCC AGA      1632
Gly Ser Thr Met Gly Ala Ala Ser Leu Thr Leu Thr Val Gln Ala Arg
    530                 535                 540

CAA TTA TTG TCT GGT ATA GTG CAG CAG CAG AAC AAT CTG CTG AGA GCT      1680
Gln Leu Leu Ser Gly Ile Val Gln Gln Gln Asn Asn Leu Leu Arg Ala
545                 550                 555                 560

ATT GAG GCG CAA CAA CAT CTG TTG CAA CTC ACA GTC TGG GGC ATC AAG      1728
Ile Glu Ala Gln Gln His Leu Leu Gln Leu Thr Val Trp Gly Ile Lys
                565                 570                 575

CAG CTC CAG GCA AGA GTC CTG GCT GTA GAA AGA TAC CTA AAG GAT CAA      1776
Gln Leu Gln Ala Arg Val Leu Ala Val Glu Arg Tyr Leu Lys Asp Gln
            580                 585                 590

CAG CTC CTG GGG ATC TGG GGT TGC TCT GGA AAA CTC ATT TGC ACC ACT      1824
Gln Leu Leu Gly Ile Trp Gly Cys Ser Gly Lys Leu Ile Cys Thr Thr
        595                 600                 605

ACT GTG CCT TGG AAT GCT AGT TGG AGT AAT AAA TCT TTG GAT CAG ATT      1872
Thr Val Pro Trp Asn Ala Ser Trp Ser Asn Lys Ser Leu Asp Gln Ile
    610                 615                 620
```

```
TGG AAT AAC ATG ACC TGG TTG GAG TGG GAC AGA GAA ATT GCC AAT TAC      1920
Trp Asn Asn Met Thr Trp Leu Glu Trp Asp Arg Glu Ile Ala Asn Tyr
625                 630                 635                 640

ACA AAC TTA ATA CAT CAC TTA ATT GAA GAA TCG CAA AAC CAG CAA GAA      1968
Thr Asn Leu Ile His His Leu Ile Glu Glu Ser Gln Asn Gln Gln Glu
                645                 650                 655

AAG AAT GAA CAA GAA TTA TTG GAA TTA GAT AAA TGG GCA AGT TTG TGG      2016
Lys Asn Glu Gln Glu Leu Leu Glu Leu Asp Lys Trp Ala Ser Leu Trp
            660                 665                 670

AGT TGG TTT GAC ATA TCA AAC TGG CTG TGG TAT ATA AAA ATA TTC ATA      2064
Ser Trp Phe Asp Ile Ser Asn Trp Leu Trp Tyr Ile Lys Ile Phe Ile
        675                 680                 685

ATG ATA GTA GCA GGC TTA GTA GGT TTA AGA ATA GTT TTT GCT GTG CTT      2112
Met Ile Val Ala Gly Leu Val Gly Leu Arg Ile Val Phe Ala Val Leu
    690                 695                 700

TCT ATA GTA AAT AGA GTT AGG CAG GGA TAC TCA CCA TTG TCA TTC CAG      2160
Ser Ile Val Asn Arg Val Arg Gln Gly Tyr Ser Pro Leu Ser Phe Gln
705                 710                 715                 720

ACC CAC TTC CCA GCT CCG AGG GGA CCC GAC AGG CCA GAC GGA ATC GAA      2208
Thr His Phe Pro Ala Pro Arg Gly Pro Asp Arg Pro Asp Gly Ile Glu
                725                 730                 735

GGA GAA GGT GGA GAG AGA GAC AGA GAC AGA TCC GTG CGA TTA GTG GAT      2256
Gly Glu Gly Gly Glu Arg Asp Arg Asp Arg Ser Val Arg Leu Val Asp
            740                 745                 750

GGA TTC TTA GCA CTT CTC TGG GAA GAC CTG CGC AAC CTG TGC CTC TTC      2304
Gly Phe Leu Ala Leu Leu Trp Glu Asp Leu Arg Asn Leu Cys Leu Phe
        755                 760                 765

AGC TAC CAC CGC TTG AGA GAC TTA CTC TTG ATT GTA ACG AGG ATT GTG      2352
Ser Tyr His Arg Leu Arg Asp Leu Leu Leu Ile Val Thr Arg Ile Val
    770                 775                 780

GAA CTT CTC GGA CGC AGG GGG TGG GAA GCC CTC AAA TAT TTG TGG AGT      2400
Glu Leu Leu Gly Arg Arg Gly Trp Glu Ala Leu Lys Tyr Leu Trp Ser
785                 790                 795                 800

CTC CTA CAG TAT TGG AGT CAG GAG CTA AAG AAT AGT GCT GTC AAC TTG      2448
Leu Leu Gln Tyr Trp Ser Gln Glu Leu Lys Asn Ser Ala Val Asn Leu
                805                 810                 815

TTC AAT ACC ACA GCT ATA GTA GTA GCT GAG GGG ACA GAT AGG ATC ATA      2496
Phe Asn Thr Thr Ala Ile Val Val Ala Glu Gly Thr Asp Arg Ile Ile
            820                 825                 830

GAA GTA GTA CAA AGA CTT TGT AGA GCT ATT CTC CAC ATA CCT AGA AGA      2544
Glu Val Val Gln Arg Leu Cys Arg Ala Ile Leu His Ile Pro Arg Arg
        835                 840                 845
```

```
ATT AGA CAG GGC TTG GAA AGA TTT TTG CTA TAA                          2577
Ile Arg Gln Gly Leu Glu Arg Phe Leu Leu  *
    850                 855
```

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 859 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
Met Arg Val Lys Glu Asn Cys Gln His Leu Trp Arg Trp Gly Trp Lys
 1               5                  10                  15

Trp Gly Ile Met Leu Leu Gly Met Leu Met Ile Cys Ser Ala Thr Glu
            20                  25                  30

Lys Leu Trp Val Thr Val Tyr Tyr Gly Val Pro Val Trp Lys Glu Thr
        35                  40                  45

Thr Thr Thr Leu Phe Cys Ala Ser Asp Ala Lys Ala Tyr Glu Glu Glu
    50                  55                  60

Val His Asn Val Trp Ala Thr His Ala Cys Val Pro Thr Asp Pro Asn
65                  70                  75                  80

Pro Gln Glu Ile Val Leu Ala Asn Val Thr Glu Asp Phe Asn Met Trp
                85                  90                  95

Lys Asn Glu Met Val Glu Gln Met His Thr Asp Ile Ile Ser Leu Trp
            100                 105                 110

Asp Glu Ser Leu Lys Pro Cys Val Lys Leu Thr Pro Leu Cys Val Thr
        115                 120                 125

Leu Asn Cys Thr Asp Leu Lys Asn Glu Thr Lys Thr Asn Ser Ser Asp
    130                 135                 140

Ala Asn Ser Asn Ser Gly Glu Ile Met Gly Asn Glu Glu Ile Lys Asn
145                 150                 155                 160

Cys Ser Phe Asn Val Ser Thr Gly Ala Pro Gly Lys Val Gln Lys Glu
                165                 170                 175

Tyr Ser Leu Phe Tyr Ala Leu Asp Ile Val Ser Ile Lys Asn Glu Asn
            180                 185                 190

Asn Ser Thr Ser His Met Leu Thr Ser Cys Asn Thr Ser Val Ser Thr
        195                 200                 205

Gln Ala Cys Pro Lys Val Ser Phe Glu Pro Ile Pro Ile His Tyr Cys
    210                 215                 220
```

```
Ala Pro Ala Gly Phe Ala Ile Leu Lys Cys Asn Asp Lys Lys Phe Asn
225                 230                 235                 240

Gly Thr Gly Pro Cys Asn Asn Val Ser Thr Val Gln Cys Thr His Gly
                245                 250                 255

Ile Arg Pro Val Val Ser Thr Gln Leu Leu Leu Asn Gly Ser Val Ala
            260                 265                 270

Glu Glu Glu Val Val Leu Arg Ser Ala Asn Phe Ser Asp Asn Ala Lys
        275                 280                 285

Thr Ile Ile Val Gln Leu Asn His Ser Val Glu Ile Thr Cys Thr Arg
    290                 295                 300

Pro Asn Tyr Asn Glu Thr Lys Lys Ile Arg Ile His Arg Gly Tyr Gly
305                 310                 315                 320

Arg Ser Phe Val Thr Val Arg Lys Leu Gly Asp Arg Lys Gln Ala His
                325                 330                 335

Cys Thr Met Asn Arg Thr Lys Trp Asp Asn Ala Leu Lys Gln Ile Ala
            340                 345                 350

Ser Lys Leu Arg Glu Gln Phe Asn Lys Thr Ala Ile Ile Phe Asn Arg
        355                 360                 365

Ser Ser Gly Gly Asp Leu Glu Ile Glu Met His Ser Phe Asn Cys Gly
    370                 375                 380

Gly Glu Leu Phe Tyr Cys Asn Thr Thr Lys Leu Phe Asn Ser Thr Trp
385                 390                 395                 400

Asn Glu Thr Thr Glu Ser Asn Gly Lys Gly Glu Asn Ile Thr Leu Pro
                405                 410                 415

Cys Arg Ile Arg Gln Phe Val Asn Met Trp Gln Lys Val Gly Lys Ala
            420                 425                 430

Met Tyr Ala Pro Pro Ser Asp Gly Gln Ile Arg Cys Thr Ser Asn Ile
        435                 440                 445

Thr Gly Leu Leu Leu Thr Arg Asp Gly Gly His Asn Asp Asn Asn Thr
    450                 455                 460

Asn Asn Glu Thr Phe Arg Pro Gly Arg Gly Asp Met Arg Asp Asn Trp
465                 470                 475                 480

Arg Ser Glu Leu Tyr Lys Tyr Lys Val Ile Lys Ile Glu Pro Leu Gly
                485                 490                 495

Val Ala Pro Thr Lys Ala Lys Arg Arg Val Val Gln Arg Glu Lys Arg
            500                 505                 510

Ala Val Gly Met Val Gly Ala Met Phe Leu Gly Phe Leu Gly Ala Ala
        515                 520                 525
```

```
Gly Ser Thr Met Gly Ala Ala Ser Leu Thr Leu Thr Val Gln Ala Arg
    530                 535                 540

Gln Leu Leu Ser Gly Ile Val Gln Gln Gln Asn Asn Leu Leu Arg Ala
545                 550                 555                 560

Ile Glu Ala Gln Gln His Leu Leu Gln Leu Thr Val Trp Gly Ile Lys
                565                 570                 575

Gln Leu Gln Ala Arg Val Leu Ala Val Glu Arg Tyr Leu Lys Asp Gln
            580                 585                 590

Gln Leu Leu Gly Ile Trp Gly Cys Ser Gly Lys Leu Ile Cys Thr Thr
        595                 600                 605

Thr Val Pro Trp Asn Ala Ser Trp Ser Asn Lys Ser Leu Asp Gln Ile
    610                 615                 620

Trp Asn Asn Met Thr Trp Leu Glu Trp Asp Arg Glu Ile Ala Asn Tyr
625                 630                 635                 640

Thr Asn Leu Ile His His Leu Ile Glu Glu Ser Gln Asn Gln Gln Glu
                645                 650                 655

Lys Asn Glu Gln Glu Leu Leu Glu Leu Asp Lys Trp Ala Ser Leu Trp
            660                 665                 670

Ser Trp Phe Asp Ile Ser Asn Trp Leu Trp Tyr Ile Lys Ile Phe Ile
        675                 680                 685

Met Ile Val Ala Gly Leu Val Gly Leu Arg Ile Val Phe Ala Val Leu
    690                 695                 700

Ser Ile Val Asn Arg Val Arg Gln Gly Tyr Ser Pro Leu Ser Phe Gln
705                 710                 715                 720

Thr His Phe Pro Ala Pro Arg Gly Pro Asp Arg Pro Asp Gly Ile Glu
                725                 730                 735

Gly Glu Gly Gly Glu Arg Asp Arg Asp Arg Ser Val Arg Leu Val Asp
            740                 745                 750

Gly Phe Leu Ala Leu Leu Trp Glu Asp Leu Arg Asn Leu Cys Leu Phe
        755                 760                 765

Ser Tyr His Arg Leu Arg Asp Leu Leu Leu Ile Val Thr Arg Ile Val
    770                 775                 780

Glu Leu Leu Gly Arg Arg Gly Trp Glu Ala Leu Lys Tyr Leu Trp Ser
785                 790                 795                 800

Leu Leu Gln Tyr Trp Ser Gln Glu Leu Lys Asn Ser Ala Val Asn Leu
                805                 810                 815

Phe Asn Thr Thr Ala Ile Val Val Ala Glu Gly Thr Asp Arg Ile Ile
            820                 825                 830
```

```
Glu Val Val Gln Arg Leu Cys Arg Ala Ile Leu His Ile Pro Arg Arg
       835                 840                 845

Ile Arg Gln Gly Leu Glu Arg Phe Leu Leu  *
    850                 855
```

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 854 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```
Met Arg Ala Arg Glu Lys Glu Arg Asn Cys Gln Asn Leu Trp Lys Trp
1               5                  10                  15

Gly Ile Met Leu Leu Gly Met Leu Met Thr Cys Ser Ala Ala Glu Asp
            20                  25                  30

Leu Trp Val Thr Val Tyr Tyr Gly Val Pro Ile Trp Lys Glu Ala Thr
        35                  40                  45

Thr Thr Leu Phe Cys Ala Ser Asp Ala Lys Ala Tyr Lys Lys Glu Ala
    50                  55                  60

His Asn Ile Trp Ala Thr His Ala Cys Val Pro Thr Asp Pro Asn Pro
65                  70                  75                  80

Gln Glu Ile Glu Leu Glu Asn Val Thr Glu Asn Phe Asn Met Trp Lys
                85                  90                  95

Asn Asn Met Val Glu Gln Met His Glu Asp Ile Ile Ser Leu Trp Asp
            100                 105                 110

Gln Ser Leu Lys Pro Cys Val Lys Leu Thr Pro Leu Cys Val Thr Leu
        115                 120                 125

Asn Cys Thr Asp Leu Lys Asn Glu Thr Lys Thr Asn Ser Ser Asp Ala
    130                 135                 140

Asn Ser Asn Ser Gly Glu Ile Met Gly Asn Glu Glu Ile Lys Asn Cys
145                 150                 155                 160

Ser Phe Asn Val Ser Thr Gly Ala Pro Gly Lys Val Gln Lys Glu Tyr
                165                 170                 175

Ser Leu Phe Tyr Ala Leu Asp Asn Asn Asn Arg Thr Asn Ser Thr Asn
            180                 185                 190
```

Tyr Arg Leu Ile Asn Cys Asp Thr Ser Thr Ile Thr Gln Ala Cys Pro
195 200 205

Lys Ile Ser Phe Glu Pro Ile Pro Ile His Phe Cys Ala Pro Ala Gly
210 215 220

Phe Ala Ile Leu Lys Cys Arg Asp Lys Lys Phe Asn Gly Thr Gly Pro
225 230 235 240

Cys Ser Asn Val Ser Thr Val Gln Cys Thr His Gly Ile Arg Pro Val
245 250 255

Val Ser Thr Gln Leu Leu Leu Asn Gly Ser Leu Ala Glu Glu Glu Ile
260 265 270

Ile Ile Arg Ser Glu Asn Leu Thr Asn Asn Val Lys Thr Ile Ile Val
275 280 285

Gln Leu Asn Ala Ser Ile Val Ile Asn Cys Thr Arg Pro Tyr Lys Tyr
290 295 300

Thr Arg Gln Arg Thr Ser Ile Gly Leu Arg Gln Ser Leu Tyr Thr Ile
305 310 315 320

Thr Gly Lys Lys Lys Lys Thr Gly Tyr Ile Gly Gln Ala His Cys Lys
325 330 335

Ile Ser Arg Ala Glu Trp Asn Lys Ala Leu Gln Gln Val Ala Thr Lys
340 345 350

Leu Gly Asn Leu Leu Asn Lys Thr Thr Ile Thr Phe Lys Pro Ser Ser
355 360 365

Gly Gly Asp Pro Glu Ile Thr Ser His Met Leu Asn Cys Gly Gly Asp
370 375 380

Phe Phe Tyr Cys Asn Thr Ser Arg Leu Phe Asn Ser Thr Trp Asn Gln
385 390 395 400

Thr Asn Ser Thr Gly Phe Asn Asn Gly Thr Val Thr Leu Pro Cys Arg
405 410 415

Ile Lys Gln Ile Val Asn Leu Trp Gln Arg Val Gly Lys Ala Met Tyr
420 425 430

Ala Pro Pro Ile Glu Gly Leu Ile Lys Cys Ser Ser Asn Ile Thr Gly
435 440 445

Leu Leu Leu Thr Arg Asp Gly Gly Ala Asn Asn Ser Ser His Glu Thr
450 455 460

Ile Arg Pro Gly Gly Gly Asp Met Arg Asp Asn Trp Arg Ser Glu Leu
465 470 475 480

Tyr Lys Tyr Lys Val Val Lys Ile Glu Pro Ile Gly Val Ala Pro Thr
485 490 495

Lys Ala Arg Arg Arg Val Val Glu Arg Glu Lys Arg Ala Ile Gly Leu
500 505 510

Gly Ala Val Phe Leu Gly Phe Leu Gly Ala Ala Gly Ser Thr Met Gly
515 520 525

Ala Ala Ser Val Thr Leu Thr Val Gln Ala Arg Gln Leu Met Ser Gly
530 535 540

Ile Val His Gln Gln Asn Asn Leu Leu Arg Ala Ile Glu Ala Gln Gln
545 550 555 560

His Leu Leu Gln Leu Thr Val Trp Gly Ile Lys Gln Leu Gln Ala Arg
565 570 575

Val Leu Ala Val Glu Arg Tyr Leu Arg Asp Gln Gln Leu Leu Gly Ile
580 585 590

Trp Gly Cys Ser Gly Arg His Ile Cys Thr Thr Asn Val Pro Trp Asn
595 600 605

Ser Ser Trp Ser Asn Arg Ser Leu Asp Glu Ile Trp Gln Asn Met Thr
610 615 620

Trp Met Glu Trp Glu Arg Glu Ile Asp Asn Tyr Thr Gly Leu Ile Tyr
625 630 635 640

Ser Leu Ile Glu Glu Ser Gln Ile Gln Gln Glu Lys Asn Glu Lys Glu
645 650 655

Leu Leu Glu Leu Asp Lys Trp Ala Ser Leu Trp Asn Trp Phe Ser Ile
660 665 670

Thr Lys Trp Leu Trp Tyr Ile Lys Leu Phe Ile Met Ile Val Gly Gly
675 680 685

Leu Ile Gly Leu Arg Ile Val Phe Ala Val Leu Ser Val Val Asn Arg
690 695 700

Val Arg Gln Gly Tyr Ser Pro Leu Ser Phe Gln Thr Leu Leu Pro Val
705 710 715 720

Pro Arg Gly Pro Asp Arg Pro Glu Glu Ile Glu Glu Glu Gly Gly Glu
725 730 735

Arg Gly Arg Asp Arg Ser Ile Arg Leu Val Asn Gly Leu Phe Ala Leu
740 745 750

Phe Trp Asp Asp Leu Arg Asn Leu Cys Leu Phe Ser Tyr His Arg Leu
755 760 765

Arg Asp Ser Ile Leu Ile Ala Ala Arg Ile Val Glu Leu Leu Gly Arg
770 775 780

Arg Gly Trp Glu Ala Leu Lys Tyr Leu Trp Asn Leu Leu Gln Tyr Trp
785 790 795 800

Ser Gln Glu Leu Arg Asn Ser Ala Ser Ser Leu Leu Asp Thr Ile Ala
805 810 815

Ile Ala Val Ala Glu Arg Thr Asp Arg Val Ile Glu Val Val Gln Arg
820 825 830

Ala Cys Arg Ala Ile Leu Asn Val Pro Arg Arg Ile Arg Gln Gly Leu
835 840 845

Glu Arg Leu Leu Leu Xaa
850

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 862 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met Arg Val Lys Glu Lys Tyr Gln His Leu Trp Arg Trp Gly Trp Lys
1 5 10 15

Trp Gly Thr Met Leu Leu Gly Ile Leu Met Ile Cys Ser Ala Thr Glu
20 25 30

Lys Leu Trp Val Thr Val Tyr Tyr Gly Val Pro Val Trp Lys Glu Ala
35 40 45

Thr Thr Thr Leu Phe Cys Ala Ser Asp Ala Lys Ala Tyr Asp Thr Glu
50 55 60

Val His Asn Val Trp Ala Thr His Ala Cys Val Pro Thr Asp Pro Asn
65 70 75 80

Pro Gln Glu Val Val Leu Val Asn Val Thr Glu Asn Phe Asn Met Trp
85 90 95

Lys Asn Asp Met Val Glu Gln Met His Glu Asp Ile Ile Ser Leu Trp
100 105 110

Asp Gln Ser Leu Lys Pro Cys Val Lys Leu Thr Pro Leu Cys Val Ser
115 120 125

Leu Lys Cys Thr Asp Leu Gly Asn Ala Thr Asn Thr Asn Ser Ser Asn
130 135 140

Thr Asn Ser Ser Ser Gly Glu Met Met Met Glu Lys Gly Glu Ile Lys
145 150 155 160

Asn Cys Ser Phe Asn Ile Ser Thr Ser Ile Arg Gly Lys Val Gln Lys
165 170 175

Glu Tyr Ala Phe Phe Tyr Lys Leu Asp Ile Ile Pro Ile Asp Asn Asp
180 185 190

Thr Thr Ser Tyr Thr Leu Thr Ser Cys Asn Thr Ser Val Ile Thr Gln
195 200 205

Ala Cys Pro Lys Val Ser Phe Glu Pro Ile Pro Ile His Tyr Cys Ala
210 215 220

Pro Ala Gly Phe Ala Ile Leu Lys Cys Asn Asn Lys Thr Phe Asn Gly
225 230 235 240

Thr Gly Pro Cys Thr Asn Val Ser Thr Val Gln Cys Thr His Gly Ile
245 250 255

Arg Pro Val Val Ser Thr Gln Leu Leu Leu Asn Gly Ser Leu Ala Glu
260 265 270

Glu Glu Val Val Ile Arg Ser Ala Asn Phe Thr Asp Asn Ala Lys Thr
275 280 285

Ile Ile Val Gln Leu Asn Gln Ser Val Glu Ile Asn Cys Thr Arg Pro
290 295 300

Asn Asn Asn Thr Arg Lys Ser Ile Arg Ile Gln Arg Gly Pro Gly Arg
305 310 315 320

Ala Phe Val Thr Ile Gly Lys Ile Gly Asn Met Arg Gln Ala His Cys
325 330 335

Asn Ile Ser Arg Ala Lys Trp Asn Ala Thr Leu Lys Gln Ile Ala Ser
340 345 350

Lys Leu Arg Glu Gln Phe Gly Asn Asn Lys Thr Ile Ile Phe Lys Gln
355 360 365

Ser Ser Gly Gly Asp Pro Glu Ile Val Thr His Ser Phe Asn Cys Gly
370 375 380

Gly Glu Phe Phe Tyr Cys Asn Ser Thr Gln Leu Phe Asn Ser Thr Trp
385 390 395 400

Phe Asn Ser Thr Trp Ser Thr Glu Gly Ser Asn Asn Thr Glu Gly Ser
405 410 415

Asp Thr Ile Thr Leu Pro Cys Arg Ile Lys Gln Phe Ile Asn Met Trp
420 425 430

Gln Glu Val Gly Lys Ala Met Tyr Ala Pro Pro Ile Ser Gly Gln Ile
435 440 445

Arg Cys Ser Ser Asn Ile Thr Gly Leu Leu Leu Thr Arg Asp Gly Gly
450 455 460

Asn Asn Asn Asn Gly Ser Glu Ile Phe Arg Pro Gly Gly Gly Asp Met
465 470 475 480

Arg Asp Asn Trp Arg Ser Glu Leu Tyr Lys Tyr Lys Val Val Lys Ile
485 490 495

Glu Pro Leu Gly Val Ala Pro Thr Lys Ala Lys Arg Arg Val Val Gln
500 505 510

Arg Glu Lys Arg Ala Val Gly Ile Gly Ala Leu Phe Leu Gly Phe Leu
515 520 525

Gly Ala Ala Gly Ser Thr Met Gly Ala Arg Ser Met Thr Leu Thr Val
530 535 540

Gln Ala Arg Gln Leu Leu Ser Gly Ile Val Gln Gln Gln Asn Asn Leu
545 550 555 560

Leu Arg Ala Ile Glu Ala Gln Gln His Leu Leu Gln Leu Thr Val Trp
565 570 575

Gly Ile Lys Gln Leu Gln Ala Arg Ile Leu Ala Val Glu Arg Tyr Leu
580 585 590

Lys Asp Gln Gln Leu Leu Gly Ile Trp Gly Cys Ser Gly Lys Leu Ile
595 600 605

Cys Thr Thr Ala Val Pro Trp Asn Ala Ser Trp Ser Asn Lys Ser Leu
610 615 620

Glu Gln Ile Trp Asn Asn Met Thr Trp Met Glu Trp Asp Arg Glu Ile
625 630 635 640

Asn Asn Tyr Thr Ser Leu Ile His Ser Leu Ile Glu Glu Ser Gln Asn
645 650 655

Gln Gln Glu Lys Asn Glu Gln Glu Leu Leu Glu Leu Asp Lys Trp Ala
660 665 670

Ser Leu Trp Asn Trp Phe Asn Ile Thr Asn Trp Leu Trp Tyr Ile Lys
675 680 685

Ile Phe Ile Met Ile Val Gly Gly Leu Val Gly Leu Arg Ile Val Phe
690 695 700

Ala Val Leu Ser Ile Val Asn Arg Val Arg Gln Gly Tyr Ser Pro Leu
705 710 715 720

Ser Phe Gln Thr His Leu Pro Thr Pro Arg Gly Pro Asp Arg Pro Glu
725 730 735

Gly Ile Glu Glu Glu Gly Gly Glu Arg Asp Arg Asp Arg Ser Ile Arg
740 745 750

Leu Val Asn Gly Ser Leu Ala Leu Ile Trp Asp Asp Leu Arg Ser Leu
755 760 765

Cys Leu Phe Ser Tyr His Arg Leu Arg Asp Leu Leu Leu Ile Val Thr
770 775 780

Arg Ile Val Glu Leu Leu Gly Arg Arg Gly Trp Glu Ala Leu Lys Tyr
785 790 795 800

Trp Trp Asn Leu Leu Gln Tyr Trp Ser Gln Glu Leu Lys Asn Ser Ala
805 810 815

Val Ser Leu Leu Asn Ala Thr Ala Ile Ala Val Ala Glu Gly Thr Asp
820 825 830

Arg Val Ile Glu Val Val Gln Gly Ala Cys Arg Ala Ile Arg His Ile
835 840 845

Pro Arg Arg Ile Arg Gln Gly Leu Glu Arg Ile Leu Leu Xaa
850 855 860

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 856 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Met Lys Val Lys Gly Thr Arg Arg Asn Tyr Gln His Leu Trp Arg Trp
1 5 10 15

Gly Thr Leu Leu Leu Gly Met Leu Met Ile Cys Ser Ala Thr Glu Lys
20 25 30

Leu Trp Val Thr Val Tyr Tyr Gly Val Pro Val Trp Lys Glu Ala Thr
35 40 45

Thr Thr Leu Phe Cys Ala Ser Asp Ala Arg Ala Tyr Asp Thr Glu Val
50 55 60

His Asn Val Trp Ala Thr His Ala Cys Val Pro Thr Asp Pro Asn Pro
65 70 75 80

Gln Glu Val Val Leu Gly Asn Val Thr Glu Asn Phe Asn Met Trp Lys
85 90 95

Asn Asn Met Val Glu Gln Met Gln Glu Asp Ile Ile Ser Leu Trp Asp
100 105 110

Gln Ser Leu Lys Pro Cys Val Lys Leu Thr Pro Leu Cys Val Thr Leu
115 120 125

Asn Cys Thr Asp Leu Gly Lys Ala Thr Asn Thr Asn Ser Ser Asn Trp
130 135 140

Lys Glu Glu Ile Lys Gly Glu Ile Lys Asn Cys Ser Phe Asn Ile Thr
145 150 155 160

Thr Ser Ile Arg Asp Lys Ile Gln Lys Glu Asn Ala Leu Phe Arg Asn
165 170 175

Leu Asp Val Val Pro Ile Asp Asn Ala Ser Thr Thr Thr Asn Tyr Thr
180 185 190

Asn Tyr Arg Leu Ile His Cys Asn Arg Ser Val Ile Thr Gln Ala Cys
195 200 205

Pro Lys Val Ser Phe Glu Pro Ile Pro Ile His Tyr Cys Thr Pro Ala
210 215 220

Gly Phe Ala Ile Leu Lys Cys Asn Asn Lys Thr Phe Asn Gly Lys Gly
225 230 235 240

Pro Cys Thr Asn Val Ser Thr Val Gln Cys Thr His Gly Ile Arg Pro
245 250 255

Ile Val Ser Thr Gln Leu Leu Leu Asn Gly Ser Leu Ala Glu Glu Glu
260 265 270

Val Val Ile Arg Ser Asp Asn Phe Thr Asn Asn Ala Lys Thr Ile Ile
275 280 285

Val Gln Leu Asn Glu Ser Val Ala Ile Asn Cys Thr Arg Pro Asn Asn
290 295 300

Asn Thr Arg Lys Ser Ile Tyr Ile Gly Pro Gly Arg Ala Phe His Thr
305 310 315 320

Thr Gly Arg Ile Ile Gly Asp Ile Arg Lys Ala His Cys Asn Ile Ser
325 330 335

Arg Ala Gln Trp Asn Asn Thr Leu Glu Gln Ile Val Lys Lys Leu Arg
340 345 350

Glu Gln Phe Gly Asn Asn Lys Thr Ile Val Phe Asn Gln Ser Ser Gly
355 360 365

Gly Asp Pro Glu Ile Val Met His Ser Phe Asn Cys Arg Gly Glu Phe
370 375 380

Phe Tyr Cys Asn Thr Thr Gln Leu Phe Asn Asn Thr Trp Arg Leu Asn
385 390 395 400

His Thr Glu Gly Thr Lys Gly Asn Asp Thr Ile Ile Leu Pro Cys Arg
405 410 415

Ile Lys Gln Ile Ile Asn Met Trp Gln Glu Val Gly Lys Ala Met Tyr
420 425 430

Ala Pro Pro Ile Gly Gly Gln Ile Ser Cys Ser Ser Asn Ile Thr Gly
435 440 445

Leu Leu Leu Thr Arg Asp Gly Gly Thr Asn Val Thr Asn Asp Thr Glu
450 455 460

Val Phe Arg Pro Gly Gly Gly Asp Met Arg Asp Asn Trp Arg Ser Glu
465 470 475 480

Leu Tyr Lys Tyr Lys Val Ile Lys Ile Glu Pro Leu Gly Ile Ala Pro
485 490 495

Thr Lys Ala Lys Arg Arg Val Val Gln Arg Glu Lys Arg Ala Val Gly
500 505 510

Ile Val Gly Ala Met Phe Leu Gly Phe Leu Gly Ala Ala Gly Ser Thr
515 520 525

Met Gly Ala Val Ser Leu Thr Leu Thr Val Gln Ala Arg Gln Leu Leu
530 535 540

Ser Gly Ile Val Gln Gln Gln Asn Asn Leu Leu Arg Ala Ile Glu Ala
545 550 555 560

Gln Gln His Leu Leu Gln Leu Thr Val Trp Gly Ile Lys Gln Leu Gln
565 570 575

Ala Arg Val Leu Ala Val Glu Arg Tyr Leu Arg Asp Gln Gln Leu Leu
580 585 590

Gly Ile Trp Gly Cys Ser Gly Lys Leu Ile Cys Thr Thr Ala Val Pro
595 600 605

Trp Asn Ala Ser Trp Ser Asn Lys Ser Leu Glu Asp Ile Trp Asp Asn
610 615 620

Met Thr Trp Met Gln Trp Glu Arg Glu Ile Asp Asn Tyr Thr Asn Thr
625 630 635 640

Ile Tyr Thr Leu Leu Glu Glu Ser Gln Asn Gln Gln Glu Lys Asn Glu
645 650 655

Gln Glu Leu Leu Glu Leu Asp Lys Trp Ala Ser Leu Trp Asn Trp Phe
660 665 670

Ser Ile Thr Asn Trp Leu Trp Tyr Ile Lys Ile Phe Ile Met Ile Val
675 680 685

Gly Gly Leu Val Gly Leu Arg Ile Val Phe Ala Val Leu Ser Ile Val
690 695 700

Asn Arg Val Arg Gln Gly Tyr Ser Pro Leu Ser Phe Gln Thr Arg Leu
705 710 715 720

Pro Val Pro Arg Gly Pro Asp Arg Pro Asp Gly Ile Glu Glu Glu Gly
725 730 735

Gly Glu Arg Asp Arg Asp Arg Ser Val Arg Leu Val Asp Gly Phe Leu
740 745 750

Ala Leu Ile Trp Glu Asp Leu Arg Ser Leu Cys Leu Phe Ser Tyr Arg
755 760 765

Arg Leu Arg Asp Leu Leu Leu Ile Ala Ala Arg Thr Val Glu Ile Leu
770 775 780

Gly His Arg Gly Trp Glu Ala Leu Lys Tyr Trp Trp Ser Leu Leu Gln
785 790 795 800

Tyr Trp Ile Gln Glu Leu Lys Asn Ser Ala Val Ser Trp Leu Asn Ala
805 810 815

Thr Ala Ile Ala Val Thr Glu Gly Thr Asp Arg Val Ile Glu Val Ala
820 825 830

Gln Arg Ala Tyr Arg Ala Ile Leu His Ile His Arg Arg Ile Arg Gln
835 840 845

Gly Leu Glu Arg Leu Leu Leu Xaa
850 855

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 854 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(ix) FEATURE:
(A) NAME/KEY: Region
(B) LOCATION: 1..854
(D) OTHER INFORMATION: /note= "Xaa indicates residues which are not specified in the consensus sequence."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Met Arg Xaa Xaa Glu Asn Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Trp Lys
1 5 10 15

Trp Gly Ile Met Leu Leu Gly Met Leu Met Xaa Ser Ala Xaa Glu Xaa
20 25 30

Leu Trp Val Thr Val Tyr Tyr Gly Val Pro Xaa Lys Glu Thr Thr Thr
35 40 45

Leu Phe Cys Ala Ser Asp Ala Lys Ala Tyr Xaa Xaa Glu Xaa His Asn
50 55 60

Xaa Trp Ala Thr His Ala Cys Val Pro Thr Asp Pro Asn Pro Gln Glu
65 70 75 80

Ile Xaa Leu Xaa Asn Val Thr Glu Xaa Phe Asn Met Trp Lys Asn Xaa
85 90 95

Met Val Glu Gln Met His Xaa Asp Ile Ile Ser Leu Trp Asp Xaa Ser
100 105 110

Leu Lys Pro Cys Val Lys Leu Thr Pro Leu Cys Val Thr Leu Asn Cys
115 120 125

Thr Asp Leu Lys Asn Glu Thr Xaa Xaa Asn Ser Ser Xaa Xaa Asn Xaa
130 135 140

Xaa Xaa Xaa Glu Ile Met Gly Xaa Glu Xaa Xaa Lys Asn Cys Ser Phe
145 150 155 160

Asn Val Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Tyr Xaa Leu
165 170 175

Phe Xaa Xaa Xaa Xaa Ile Val Xaa Xaa Xaa Asn Xaa Xaa Asn Ser Thr
180 185 190

Xaa Xaa Xaa Leu Xaa Xaa Cys Xaa Thr Ser Xaa Xaa Thr Gln Ala Cys
195 200 205

Pro Lys Xaa Ser Phe Glu Pro Ile Pro Ile His Xaa Cys Ala Pro Ala
210 215 220

Gly Phe Ala Ile Leu Lys Cys Xaa Asp Lys Lys Phe Asn Gly Thr Gly
225 230 235 240

Pro Cys Xaa Asn Val Ser Thr Val Gln Cys Thr His Gly Ile Arg Pro
245 250 255

Val Val Ser Thr Gln Leu Leu Leu Asn Gly Ser Xaa Ala Glu Glu Glu
260 265 270

Xaa Xaa Xaa Arg Ser Xaa Asn Xaa Xaa Xaa Asn Xaa Lys Thr Ile Ile
275 280 285

Val Gln Leu Asn Xaa Ser Xaa Xaa Ile Xaa Cys Thr Arg Pro Xaa Xaa
290 295 300

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Ile Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
305 310 315 320

Xaa Thr Gly Lys Xaa Xaa Lys Xaa Gly Xaa Xaa Xaa Gln Ala His Cys
325 330 335

Xaa Xaa Xaa Arg Xaa Xaa Trp Xaa Xaa Ala Leu Xaa Gln Xaa Ala Xaa
340 345 350

Lys Leu Xaa Xaa Xaa Xaa Asn Lys Thr Xaa Ile Xaa Phe Xaa Xaa Ser
355 360 365

Ser Gly Gly Asp Xaa Glu Ile Xaa Xaa His Xaa Xaa Asn Cys Gly Gly
370 375 380

Xaa Xaa Phe Tyr Cys Asn Thr Xaa Xaa Leu Phe Asn Ser Thr Trp Asn
385 390 395 400

Xaa Thr Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Thr Leu Pro Cys
405 410 415

Arg Ile Xaa Gln Xaa Val Asn Xaa Trp Gln Xaa Val Gly Lys Ala Met
420 425 430

Tyr Ala Pro Pro Xaa Xaa Gly Xaa Ile Xaa Cys Ser Asn Ile Thr Gly
435 440 445

Leu Leu Leu Thr Arg Asp Gly Gly His Asn Xaa Asn Asn Xaa Xaa Xaa
450 455 460

Glu Thr Xaa Arg Pro Gly Xaa Gly Asp Met Arg Asp Asn Trp Arg Ser
465 470 475 480

Glu Leu Tyr Lys Tyr Lys Val Xaa Lys Ile Glu Pro Xaa Gly Val Ala
485 490 495

Pro Thr Lys Ala Xaa Arg Arg Val Val Xaa Arg Glu Lys Arg Ala Xaa
500 505 510

Met Gly Ala Xaa Phe Leu Gly Phe Leu Gly Ala Ala Gly Ser Thr Met
515 520 525

Gly Ala Ala Ser Xaa Leu Thr Val Gln Ala Arg Gln Leu Xaa Ser Gly
530 535 540

Ile Val Xaa Gln Gln Asn Asn Leu Leu Arg Ala Ile Glu Ala Gln Gln
545 550 555 560

His Leu Leu Gln Leu Thr Val Trp Gly Ile Lys Gln Leu Gln Ala Arg
565 570 575

Val Leu Ala Val Glu Arg Tyr Leu Xaa Asp Gln Gln Leu Leu Gly Ile
580 585 590

Trp Gly Cys Ser Gly Xaa Xaa Ile Cys Thr Thr Xaa Val Pro Trp Asn
595 600 605

Xaa Ser Trp Ser Asn Xaa Ser Leu Asp Xaa Ile Trp Xaa Asn Met Thr
610 615 620

Trp Xaa Glu Trp Xaa Arg Glu Ile Xaa Asn Tyr Thr Xaa Leu Ile Xaa
625 630 635 640

Xaa Leu Ile Glu Glu Ser Gln Xaa Gln Gln Glu Lys Asn Glu Xaa Glu
645 650 655

Leu Leu Glu Leu Asp Lys Trp Ala Ser Leu Trp Xaa Trp Phe Xaa Ile
660 665 670

Xaa Xaa Trp Leu Trp Tyr Ile Lys Xaa Phe Ile Met Ile Val Xaa Gly
675 680 685

Leu Xaa Gly Leu Arg Ile Val Phe Ala Val Leu Ser Xaa Val Asn Arg
690 695 700

Val Arg Gln Gly Tyr Ser Pro Leu Ser Phe Gln Thr Xaa Xaa Pro Xaa
705 710 715 720

Pro Arg Gly Pro Asp Arg Pro Xaa Xaa Ile Glu Xaa Glu Gly Gly Glu
725 730 735

Arg Xaa Arg Asp Arg Ser Xaa Arg Leu Val Xaa Gly Xaa Xaa Ala Leu
740 745 750

Xaa Trp Xaa Asp Leu Arg Asn Leu Cys Leu Phe Ser Tyr His Arg Leu
755 760 765

Arg Asp Xaa Xaa Leu Ile Xaa Xaa Arg Ile Val Glu Leu Leu Gly Arg
770 775 780

Arg Gly Trp Glu Ala Leu Lys Tyr Leu Trp Xaa Leu Leu Gln Tyr Trp
785 790 795 800

Ser Gln Glu Leu Xaa Asn Ser Ala Xaa Xaa Leu Xaa Xaa Thr Xaa Ala
805 810 815

Ile Xaa Val Ala Glu Xaa Thr Asp Arg Xaa Ile Glu Val Val Gln Arg
820 825 830

Xaa Cys Arg Ala Ile Leu Xaa Xaa Pro Arg Arg Ile Arg Gln Gly Leu
835 840 845

Glu Arg Xaa Leu Leu Xaa
850

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 866 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(ix) FEATURE:
(A) NAME/KEY: Region
(B) LOCATION: 1..866
(D) OTHER INFORMATION: /note= "Xaa residues are not specified in this consensus sequence."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Met Arg Val Lys Glu Xaa Xaa Gln His Leu Trp Arg Trp Gly Trp Lys
1 5 10 15

Trp Gly Xaa Met Leu Leu Gly Xaa Leu Met Ile Cys Ser Ala Thr Glu
20 25 30

Lys Leu Trp Val Thr Val Tyr Tyr Gly Val Pro Val Trp Lys Glu Xaa
35 40 45

Thr Thr Thr Leu Phe Cys Ala Ser Asp Ala Lys Ala Tyr Xaa Xaa Glu
50 55 60

Val His Asn Val Trp Ala Thr His Ala Cys Val Pro Thr Asp Pro Asn
65 70 75 80

Pro Gln Glu Xaa Val Leu Xaa Asn Val Thr Glu Xaa Phe Asn Met Trp
85 90 95

Lys Asn Xaa Met Val Glu Gln Met His Xaa Asp Ile Ile Ser Leu Trp
100 105 110

Asp Xaa Ser Leu Lys Pro Cys Val Lys Leu Thr Pro Leu Cys Val Xaa
115 120 125

Leu Xaa Cys Thr Asp Leu Xaa Asn Xaa Thr Xaa Thr Asn Ser Ser Xaa
130 135 140

Xaa Asn Ser Xaa Ser Gly Glu Xaa Met Met Xaa Xaa Xaa Glu Ile Lys
145 150 155 160

Asn Cys Ser Phe Asn Xaa Ser Thr Xaa Xaa Xaa Gly Lys Val Gln Lys
165 170 175

Glu Tyr Xaa Xaa Phe Tyr Xaa Leu Asp Ile Val Ser Ile Xaa Xaa Xaa
180 185 190

Asn Xaa Xaa Thr Ser Xaa Xaa Leu Thr Ser Cys Asn Thr Ser Val Xaa
195 200 205

Thr Gln Ala Cys Pro Lys Val Ser Phe Glu Pro Ile Pro Ile His Tyr
210 215 220

Cys Ala Pro Ala Gly Phe Ala Ile Leu Lys Cys Asn Xaa Lys Xaa Phe
225 230 235 240

Asn Gly Thr Gly Pro Cys Xaa Asn Val Ser Thr Val Gln Cys Thr His
245 250 255

Gly Ile Arg Pro Val Val Ser Thr Gln Leu Leu Leu Asn Gly Ser Xaa
260 265 270

Ala Glu Glu Glu Val Val Xaa Arg Ser Ala Asn Phe Xaa Asp Asn Ala
275 280 285

Lys Thr Ile Ile Val Gln Leu Asn Xaa Ser Val Glu Ile Xaa Cys Thr
290 295 300

```
Arg Pro Asn Xaa Asn Xaa Xaa Lys Xaa Ile Arg Ile Xaa Arg Gly Xaa
305                 310                 315                 320

Gly Arg Xaa Phe Val Thr Xaa Xaa Lys Xaa Gly Xaa Xaa Xaa Gln Ala
                325                 330                 335

His Cys Xaa Xaa Xaa Arg Xaa Lys Trp Xaa Xaa Xaa Leu Lys Gln Ile
            340                 345                 350

Ala Ser Lys Leu Arg Glu Gln Phe Gly Asn Xaa Xaa Xaa Ile Ile Phe
        355                 360                 365

Xaa Xaa Ser Ser Gly Gly Asp Xaa Glu Ile Xaa Xaa His Ser Phe Asn
    370                 375                 380

Cys Gly Gly Glu Xaa Phe Tyr Cys Asn Xaa Thr Xaa Leu Phe Asn Ser
385                 390                 395                 400

Thr Trp Phe Asn Xaa Thr Trp Ser Thr Glu Gly Ser Asn Asn Xaa Xaa
                405                 410                 415

Gly Ser Xaa Xaa Ile Thr Leu Pro Cys Arg Ile Xaa Gln Phe Xaa Asn
            420                 425                 430

Met Trp Gln Xaa Val Gly Lys Ala Met Tyr Ala Pro Pro Xaa Xaa Gly
        435                 440                 445

Gln Ile Arg Cys Xaa Ser Asn Ile Thr Gly Leu Leu Leu Thr Arg Asp
    450                 455                 460

Gly Gly His Asn Asp Asn Asn Xaa Xaa Xaa Glu Xaa Phe Arg Pro Gly
465                 470                 475                 480

Xaa Gly Asp Met Arg Asp Asn Trp Arg Ser Glu Leu Tyr Lys Tyr Lys
                485                 490                 495

Val Xaa Lys Ile Glu Pro Leu Gly Val Ala Pro Thr Lys Ala Lys Arg
            500                 505                 510

Arg Val Val Gln Arg Glu Lys Arg Ala Val Gly Met Xaa Gly Ala Xaa
        515                 520                 525

Phe Leu Gly Phe Leu Gly Ala Ala Gly Ser Thr Met Gly Ala Xaa Ser
    530                 535                 540

Xaa Thr Leu Thr Val Gln Ala Arg Gln Leu Leu Ser Gly Ile Val Gln
545                 550                 555                 560

Gln Gln Asn Asn Leu Leu Arg Ala Ile Glu Ala Gln Gln His Leu Leu
                565                 570                 575

Gln Leu Thr Val Trp Gly Ile Lys Gln Leu Gln Ala Arg Xaa Leu Ala
            580                 585                 590

Val Glu Arg Tyr Leu Lys Asp Gln Gln Leu Leu Gly Ile Trp Gly Cys
        595                 600                 605
```

```
Ser Gly Lys Leu Ile Cys Thr Thr Xaa Val Pro Trp Asn Ala Ser Trp
    610                 615                 620

Ser Asn Lys Ser Leu Xaa Gln Ile Trp Asn Asn Met Thr Trp Xaa Glu
625                 630                 635                 640

Trp Asp Arg Glu Ile Xaa Asn Thr Xaa Leu Ile His Xaa Leu Ile Glu
                645                 650                 655

Glu Ser Gln Asn Gln Gln Glu Lys Asn Glu Gln Glu Leu Leu Glu Leu
            660                 665                 670

Asp Lys Trp Ala Ser Leu Trp Xaa Trp Phe Xaa Ile Xaa Asn Trp Leu
        675                 680                 685

Trp Tyr Ile Lys Ile Phe Ile Met Ile Val Xaa Gly Leu Val Gly Leu
    690                 695                 700

Arg Ile Val Phe Ala Val Leu Ser Ile Val Asn Arg Val Arg Gln Gly
705                 710                 715                 720

Tyr Ser Pro Leu Ser Phe Gln Thr His Xaa Pro Xaa Pro Arg Gly Pro
                725                 730                 735

Asp Arg Pro Xaa Gly Ile Glu Xaa Glu Gly Gly Glu Arg Asp Arg Asp
            740                 745                 750

Arg Ser Xaa Arg Leu Val Xaa Gly Xaa Leu Ala Leu Xaa Trp Xaa Asp
        755                 760                 765

Leu Arg Xaa Leu Cys Leu Phe Ser Tyr His Arg Leu Arg Asp Leu Leu
    770                 775                 780

Leu Ile Val Thr Arg Ile Val Glu Leu Leu Gly Arg Arg Gly Trp Glu
785                 790                 795                 800

Ala Leu Lys Tyr Xaa Trp Xaa Leu Leu Gln Tyr Trp Ser Gln Glu Leu
                805                 810                 815

Lys Asn Ser Ala Val Xaa Leu Xaa Asn Xaa Thr Ala Ile Xaa Val Ala
            820                 825                 830

Glu Gly Thr Asp Arg Xaa Ile Glu Val Val Gln Xaa Xaa Cys Arg Ala
        835                 840                 845

Ile Xaa His Ile Pro Arg Arg Ile Arg Gln Gly Leu Glu Arg Xaa Leu
    850                 855                 860

Leu Xaa
865
```

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 836 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant

(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(ix) FEATURE:
(A) NAME/KEY: Region
(B) LOCATION: 1..836
(D) OTHER INFORMATION: /note= "Xaa residues are not specified in this consensus sequence."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Met Val Lys Glu Trp Trp Gly Leu Leu Gly Met Leu Met Ile Cys Ser
1 5 10 15

Ala Thr Glu Lys Leu Trp Val Thr Val Tyr Tyr Gly Val Pro Val Trp
20 25 30

Lys Glu Thr Thr Thr Leu Phe Cys Ala Ser Asp Ala Ala Tyr Glu Val
35 40 45

His Asn Val Trp Ala Thr His Ala Cys Val Pro Thr Asp Pro Asn Pro
50 55 60

Gln Glu Val Leu Asn Val Thr Glu Phe Asn Met Trp Lys Asn Met Val
65 70 75 80

Glu Gln Met Asp Ile Ile Ser Leu Trp Asp Ser Leu Lys Pro Cys Val
85 90 95

Lys Leu Thr Pro Leu Cys Val Thr Leu Asn Cys Thr Asp Leu Xaa Xaa
100 105 110

Xaa Thr Xaa Thr Asn Ser Ser Asp Ala Asn Xaa Xaa Xaa Xaa Xaa Xaa
115 120 125

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
130 135 140

Xaa Xaa Xaa Lys Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
145 150 155 160

Ile Xaa Xaa Xaa Xaa Xaa Xaa Xaa Asn Xaa Thr Xaa Xaa Xaa Leu Xaa
165 170 175

Xaa Cys Asn Xaa Ser Val Xaa Thr Gln Ala Cys Pro Lys Val Ser Phe
180 185 190

Glu Pro Ile Pro Ile His Tyr Cys Xaa Pro Ala Gly Phe Ala Ile Leu
195 200 205

Lys Cys Asn Xaa Lys Xaa Phe Asn Gly Xaa Gly Pro Cys Xaa Asn Val
210 215 220

Ser Thr Val Gln Cys Thr His Gly Ile Arg Pro Xaa Val Ser Thr Gln
225 230 235 240

Leu Leu Leu Asn Gly Ser Xaa Ala Glu Glu Glu Val Val Xaa Arg Ser
245 250 255

Xaa Asn Phe Xaa Xaa Asn Ala Lys Thr Ile Ile Val Gln Leu Asn Xaa
260 265 270

Ser Val Xaa Ile Xaa Cys Thr Arg Pro Asn Xaa Asn Xaa Xaa Lys Xaa
275 280 285

Ile Xaa Ile Xaa Xaa Gly Tyr Xaa Xaa Xaa Xaa Xaa Thr Xaa Arg Xaa
290 295 300

Xaa Gly Asp Xaa Xaa Xaa Ala His Cys Xaa Xaa Xaa Arg Xaa Xaa Trp
305 310 315 320

Xaa Asn Xaa Leu Xaa Gln Ile Xaa Xaa Lys Leu Arg Glu Gln Phe Gly
325 330 335

Asn Xaa Xaa Xaa Ile Xaa Phe Asn Xaa Ser Ser Gly Gly Asp Xaa Glu
340 345 350

Ile Xaa Met His Ser Phe Asn Cys Xaa Gly Glu Xaa Phe Tyr Cys Asn
355 360 365

Thr Thr Xaa Leu Phe Asn Xaa Thr Trp Arg Xaa Xaa Xaa Thr Glu Xaa
370 375 380

Xaa Xaa Xaa Xaa Xaa Xaa Ile Xaa Leu Pro Cys Arg Ile Xaa Gln Xaa
385 390 395 400

Xaa Asn Met Trp Gln Xaa Val Gly Lys Ala Met Tyr Ala Pro Pro Xaa
405 410 415

Xaa Gly Gln Ile Xaa Cys Xaa Ser Asn Ile Thr Gly Leu Leu Leu Thr
420 425 430

Arg Asp Gly Gly Xaa Asn Xaa Xaa Asn Thr Xaa Xaa Glu Xaa Phe Arg
435 440 445

Pro Gly Xaa Gly Asp Met Arg Asp Asn Trp Arg Ser Glu Leu Tyr Lys
450 455 460

Tyr Lys Val Ile Lys Ile Glu Pro Leu Gly Xaa Ala Pro Thr Lys Ala
465 470 475 480

Lys Arg Arg Val Val Gln Arg Glu Lys Arg Ala Val Gly Xaa Val Gly
485 490 495

Ala Met Phe Leu Gly Phe Leu Gly Ala Ala Gly Ser Thr Met Gly Ala
500 505 510

Xaa Leu Thr Leu Thr Val Gln Ala Arg Gln Leu Leu Ser Gly Ile Val
515 520 525

Gln Gln Gln Asn Asn Leu Leu Arg Ala Ile Glu Ala Gln Gln His Leu
530 535 540

Leu Gln Leu Thr Val Trp Gly Ile Lys Gln Leu Gln Ala Arg Val Leu
545 550 555 560

Ala Val Glu Arg Tyr Leu Xaa Asp Gln Gln Leu Leu Gly Ile Trp Gly
565 570 575

Cys Ser Gly Lys Leu Ile Cys Thr Thr Xaa Val Pro Trp Asn Ala Ser
580 585 590

Trp Ser Asn Lys Ser Leu Xaa Xaa Ile Trp Xaa Asn Met Thr Trp Xaa
595 600 605

Xaa Trp Xaa Arg Glu Ile Xaa Asn Tyr Thr Asn Xaa Ile Xaa Xaa Leu
610 615 620

Xaa Glu Glu Ser Gln Asn Gln Gln Glu Lys Asn Glu Gln Glu Leu Leu
625 630 635 640

Glu Leu Asp Lys Trp Ala Ser Leu Trp Xaa Trp Phe Xaa Ile Xaa Asn
645 650 655

Trp Leu Trp Tyr Ile Lys Ile Phe Ile Met Ile Val Xaa Gly Leu Val
660 665 670

Gly Leu Arg Ile Val Phe Ala Val Leu Ser Ile Val Asn Arg Val Arg
675 680 685

Gln Gly Tyr Ser Pro Leu Ser Phe Gln Thr Xaa Xaa Pro Xaa Pro Arg
690 695 700

Gly Pro Asp Arg Pro Asp Gly Ile Glu Xaa Glu Gly Gly Glu Arg Asp
705 710 715 720

Arg Asp Arg Ser Val Arg Leu Val Asp Gly Phe Leu Ala Leu Xaa Trp
725 730 735

Glu Asp Leu Arg Xaa Leu Cys Leu Phe Ser Tyr Xaa Arg Leu Arg Asp
740 745 750

Leu Leu Leu Ile Xaa Xaa Arg Xaa Val Glu Xaa Leu Gly Xaa Arg Gly
755 760 765

Trp Glu Ala Leu Lys Tyr Xaa Trp Ser Leu Leu Gln Tyr Trp Xaa Gln
770 775 780

Glu Leu Lys Asn Ser Ala Val Xaa Xaa Xaa Asn Xaa Thr Ala Ile Xaa
785 790 795 800

Val Xaa Glu Gly Thr Asp Arg Xaa Ile Glu Val Xaa Gln Arg Xaa Xaa
805 810 815

Arg Ala Ile Leu His Ile Xaa Arg Arg Ile Arg Gln Gly Leu Glu Arg
820 825 830

Xaa Leu Leu Xaa
835

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 39 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "Oligonucleotide."

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

CCCTTCGAAG AGGATATAAT CAGTTTATGG GATCAAAGC 39

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 33 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "Oligonucleotide."

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

CCCTTCGAAC TCTTCTTCTG CTAGACTGCC ATT 33

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9193 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

GGTCTCTCTG GTTAGACCAG ATCTGAGCCT GGGAGCTCTC TGGCTAGCTA GGGAACCCAC 60

TGCTTAAGCC TCAATAAAGC TTGCCTTGAG TGCTTCAAGT AGTGTGTGCC CGTCTGTTGT 120

GTGACTCTGG TAACTAGAGA TCCCTCAGAC CCTTTTAGTC AGTGTGGAAA AATCTCTAGC 180

AGTGGCGCCC GAACAGGGAC CGGAAAGCAA AAGGGAAACC AGAGAAGCTC TCTCGACGCA 240

GGACTCGGCT TGCTGAAGCG CGCACAGCAA GAGGCGAGGG GCGGCGACTG GTGAGTACGC 300

CGATATTTTT GACTAGCGGA GGCTAGAAGG AGAGAGATGG GTGCGAGAGC GTCGGTATTA 360

AGCGGGGGAG ATTTAGATCG ATGGGAAAAA ATTCGGTTAA GGCCAGGGGG AAAGAAAAAA 420

TATATGTTAA AACATATAGT ATGGGCAAGC AGGGAGCTAG AACGATTCGC AGTCAATCCT 480

GGCCTGTTAG AAACATCAGA AGGCTGTAGA CAAGTACTGG AACAGCTACA GCCATCCCTT 540

CAGACAGGAT CAGAAGAACT TAGATCCTTA TATAATACAA TAGCAACCCT CTATGGTGTG 600

CATCAAAAAA TAGAGGTAAA AGACACCAAG GAAGCTTTAG ACAAAATAGA GGAAGAGCAA 660

AACAAAAGTA AGAAAAAAGC ACAGCAAGCA GCAGCTGACA CAGGACACAG CAGCCAGGTC 720

AGCCAAAATT ACCCTATAGT GCAGAACCTA CAGGGGCAAA TGGTACATCA GGCCATATCA 780

CCTAGAACTT TAAATGCATG GGTAAAAGTA ATAGAAGAGA AGGCTTTCAG CCCAGAAGTG 840

ATACCCATGT TTTCAGCATT ATCAGAAGGA GCCACCCCAC AAGATTTAAA CACCATGTTA 900

AACACAGTGG GGGGACATCA AGCAGCCATG CAAATGTTAA AAGAGACCAT CAATGAGGAA 960

GCTGCAGAAT GGGATAGAGT GCATCCAGTG CAGGCAGGGC TTATTGCACC AGGCCAGATG 1020

AGAGAACCAA GGGGAAGTGA CATAGCAGGA ACTACTAGTA CCCTTCAGGA ACAAATAGGA 1080

TGGATGACAA GTAATCCACC TATCCCAGTA GGAGAAATTT ATAAAAGGTG GATAATCCTG 1140

GGCTTAAATA AAATAGTAAG AATGTATAGC CCTATCAGCA TTCTAGACAT AAGACAAGGA 1200

CCAAAAGAAC CCTTTAGAGA CTATGTAGAC CGGTTCTATA AAACTCTAAG AGCCGAGCAA 1260

GCTTCACAGG AAGTAAAAAA TTGGATGACA GAAACCTTGC TGGTCCAAAA TGCGAACCCA 1320

GATTGTAAGA CTATCTTAAA AGCATTAGGA CCAGGAGCTA CACTAGAAGA AATGATGACA 1380

GCATGTCAGG GAGTGGGAGG ACCCGGCCAT AAGGCAAGAG TTTTAGCTGA AGCAATGAGC 1440

CAAGTAACAA ATTCAGCTGC CATAATGATG CAGAGAGGCA ATTTTAAGAA CCAAAGAAAG 1500

ATGGTTAAGT GTTTCAATTG TGGCAAAGAG GGGCACGTAG CCAGAAATTG CAGGGCCCCT 1560

AGAAAAAAGG GCTGTTGGAA ATGTGGAAAG GAAGGACACC AAATGAAAGA TTGTATTGAG 1620

AGCCAGGCTA ATTTTTTAGG GAAAATCTGG CCTTCCCACA AGGGAAGGCC AGGGAATTTC 1680

CTTCAGAGCA GACCAGAGCC AACAGCCCCA CCAGAAGAGA GCCTCAGGTC TGGGATAGAG 1740

ACAACAACTC CCTCTCAGAA GCAGGAGCCA ATAGACAAGG AAGTGTATCC TTTAACTTCC 1800

CTCAGATCAC TCTTTGGCAA CGACCCCTCG TCACAATAAA GATAGGGGGG CAACTAAAGG 1860

AAGCTCTATT AGATACAGGA GCAGATGATA CAGTATTAGA AGAAATGAAT TTGCCAGGAA 1920

GATGGAAGCC AAAAATGATA GGGGGCATTG GAGGTTTTAT CAAAGTAAGA CAGTATGATC 1980

AGATACCCAT AGAAATCTGT GGACATAAAG CTATAGGTAC AGTATTAGTA GGACCTACAC 2040

CTGTCAACAT AATTGGAAGA AATCTGTTGA CTCAGATTGG TTGCACTTTA AATTTTCCCA 2100

TTAGTCCTAT TGAAACTGTA CCAGTGAGAT TAAAGCCAGG AATGGATGGC CCAAAAGTTA 2160

AACAATGGCC ATTGACAGAA GAAAAAATAA AAGCATTAGT AGAAATTTGT ACAGAAATGG 2220

AAAAGGAAGG GAAAATTTCA AAAATTGGGC CTGAAAATCC ATACAATACT CCAGTATTTG 2280

CCATAAAGAA AAAAGACAGT ACTAAATGGA GAAAATTAGT AGATTTCAGA GAACTTAATA 2340

AGAGAACTCA AGACTTCTGG GAAGTTCAAT TAGGAATACC ACATCCCTCA GGTTTAAAAA 2400

AGAAAAAATC AGTAACAGTA CTAGATGTGG GTGATGCATA TTTTTCAGTT CCCTTAGATG 2460

AAGATTTCAG GAAGTATACT ACATTTACCA TACCTAGTAT AAACAATCAG ACACCAGGGA 2520

TTAGATATCA GTACAATGTG CTTCCACAGG GATGGAAAGG ATCACCAGCA ATATTCCAAA 2580

GTAGCATGAC AAAAATCTTA GAGCCTTTTA GAAAACAGAA TCCAGACATA GTTATCTATC 2640

AATACATGGA TGATTTGTAT GTAGGATCTG ACTTAGAAAT AGAGCAGCAT AGAACAAAAA 2700

TAGAGGAACT GAGACAGCAT CTGTTGAGGT GGGGATTTAC CACACCAGAC AAAAAACATC 2760

AGAAAGAACC TCCATTCCTC TGGATGGGGT ATGAACTCCA TCCTGATAAA TGGACAGTAC 2820

AGCCTATAGT GCTGCCAGAA AAAGACAGCT GGACTGTCAA TGACATACAG AAGTTAGTGG 2880

GAAAATTGAA TTGGGCAAGT CAGATTTACC CAGGGATTAA AGTAAGGCAA TTATGTAAAC 2940

TCCTTAGAGG AGCCAAAGCA CTAACAGAAG TAATACCACT AACAGAAGGA GCAGAGCTAG 3000

AACTGGCAGA AAACAGAGAG ATTCTAAAAG AACCAGTACA TGGAGTGTAT TATGACCCAT 3060

CAAAAGACTT AATAGCAGAA ATACAGAAGC AGGGGCAAGG CCAATGGACA TATCAAATTT 3120

ATCAAGAGCC ATTTAAAAAT CTGAAAACAG GAAAATATGC AAGAACGAGG GGTGCCCACA 3180

CTAATGATGT AAAACAATTA ACAGAGGCAG TGCAAAAAAT AGCCACAGAA AGCATAGTAA 3240

TATGGGGGAA GACTCCTAAA TTTAAACTGC CCATACAAAA GGAAACATGG GAAACATGGT 3300

GGACAGAGTA TTGGCAAGCC ACTTGGATTC CTGAATGGGA GTTTGTCAAT ACCCCTCCTT 3360

TAGTAAAATT ATGGTACCAG TTAGAGAAAG AACCCATAGT AGGAGCAGAA ACTTTCTATG 3420

TAGATGGGGC AGCTAACAGG GAGACTAAAT TAGGAAAAGC AGGATATGTT ACTAACAGAG 3480

GAAGACAAAA GGTTGTCACC CTGACTAACA CAACAAATCA GAAGACTGAG TTACAAGCAA 3540

TTCATCTAGC TTTACAGGAT TCAGGATCAG AAGTAAACAT AGTAACAGAC TCACAATATG 3600

CATTAGGAAT CATTCAAGCA CAACCAGATC AAAGTGAATC AGAGTTAGTC AATCAAATAA 3660

TAGAGCAGTT AATAAAAAAG GAAAAGGTCT ATCTGGCATG GGTACCAGCA CACAAAGGAA 3720

TTGGAGGAAA TGAACAAGTA GATAAATTAG TCAGTGCTGG AATCAGAAAA GTACTATTTT 3780

TAGATGGAAT AGATAAGGCC CAAGAAGAAC ATGAGAAATA TCACAATAAT TGGAGAGCAA 3840

TGGCTAGTGA TTTTAACCTG CCACCTGTAG TAGCAAAAGA AATAGTAGCC AGCTGTGATA 3900

AATGTCAGCT AAAAGGAGAA GCTATGCATG GACAAGTAGA CTGTAGTCCA GGAATATGGC 3960

AACTAGATTG TACACATTTA GAAGGAAAAG TTATCCTGGT AGCAGTTCAT GTAGCCAGTG 4020

GATATATAGA AGCAGAAGTT ATTCCAGCAG AAACAGGGCA GGAAACAGCA TACTTTCTCT 4080

TAAAATTAGC AGGAAGATGG CCAGTAAAAA CAATACATAC AGACAATGGC AGCAATTTCA 4140

CCAGTGCTGC GGTTAAGGCC GCCTGTTGGT GGGCAGGAGT CAAACAAGAA TTTGGAATTC 4200

CCTACAATCC CCAAAGTCAA GGAGTAGTAG AATCTATGAA TAAAGAATTA AAGAAAATTA 4260

TAGGACAGGT AAGAGATCAA GCTGAACATC TTAAGACAGC AGTACAAATG GCAGTATTTG 4320

TCCACAATTT TAAAAGAAAA GGGGGGATTG GGGGGTACAG TGCAGGGGAA AGAATAATAG 4380

ACATAATAGC AACAGACATA CAAACTAGAG AACTACAAAA ACAAATTACA AAAATTCAAA 4440

ATTTTCGGGT TTATTACAGG GACAGCAGAG ATCCACTTTG GAAAGGACCA GCAAAGCTCC 4500

TCTGGAAAGG TGAAGGGGCA GTAGTAATAC AAGATAATAG TGACATAAAA GTAGTGCCAA 4560

GAAGAAAAGC AAAGATCATT AGGGATTATG GAAAACAGAT GGCAGGTGAT GATTGTGTGG 4620

CAAGTAGACA GGATGAGGAT TAGAACATGG AAAAGTTTAG TAAAACATCA TGTATATGTT 4680

TCAAAGAAAG CTAGGGGATG GTTTTATAGA CATCACTATG AAAGCACTCA TCCAAAAATA 4740

AGTTCAGAAA TACACATCCC ACTAGGGGAT GCTAGATTGG TAGTAACAAC ATATTGGGGT 4800

CTGCATACAG GAGAAAGAGA ATGGCATTTG GGTCATGGAG TCTCCATAGA ATGGAGGAAA 4860

AGGAGCTATA GCACACAAGT AGACCCTGAA CTAGCAGACC AACTAATTCA TCTGTATTAC 4920

TTTGATTGTT TTTCAGACTC TGCTATAAGA AAGGCCTTAT TAGAACACAT AGTTAGCCCT 4980

AGGTGTGAAT ATCGAGCAGG ACATTCCAAG GTAGGATCTC TACAATACTT GGCACTATCA 5040

GCCTTAATAA CACCAAAGAA GATAAAGCCA CCTTTGCCTA GTGTTACGAA ACTGACAGAG 5100

GATAGATGGA ACAAGCCCCA GAAGACCAAG GGCCACAGAG GGAGCCATAC AATGAATGGA 5160

CACTAGAGCT TTTAGAGGAG CTTAAGGGAG AAGCTGTTAG ACATTTTCCT AGGCCATGGC 5220

TCCATAGCTT AGGACAACAT ATCTATGAAA CTTATGGAGA TACTTGGGCA GGAGTGGAAG 5280

CCATAATAAG AATTCTGCAA CAATTGCTGT TTATTCATTT CAGAATTGGG TGTCAACATA 5340

GCAGAATAGG CATTATTCAA CAGAGGAGAG CAAGAAGAAA TGGAGCCAGT AGATCCTAAT 5400

CTAGAGCCCT GGAAGCATCC AGGAAGTCAG CCTAAAACTG CTTGTACCAA ATGCTATTGT 5460

AAAAGGTGTT GCTTTCATTG CCAAGTTTGT TTCACAACAA AAGCCTTAGG CATCTCCTAT 5520

GGCAGGAAGA AGCGGAGACA GCGACGAAGA CCTCGTCAGG GCAGCCAGGC TCATCAAGCT 5580

TCTCTATCAG AGCAGTAAGT AGTATATGTA ATGCAACTTA TATTAATTGT AACAATAGTA 5640

GCTTTAGTAG TAACATTAAT AATAGCAATA GTTGTGTGGT CCATAGTACT CATAGAATAT 5700

AGGAAAATAT TAAGACAAAG GAAAATAGAC AAGTTAATTA ATAGACTAGT AGAAAGAGCA 5760

GAAGACAGTG GCAATGAGAG TGAAGGAGAA CTGTCAGCAC TTGTGGAGAT GGGGGTGGAA 5820

ATGGGGCATC ATGCTCCTTG GGATGTTAAT GATCTGTAGT GCTACAGAAA AATTGTGGGT 5880

CACAGTCTAT TATGGGGTAC CTGTGTGGAA GGAAACAACT ACCACTCTAT TTTGTGCATC 5940

AGATGCTAAA GCATATGAAG AAGAGGTACA TAATGTTTGG GCCACACATG CCTGTGTACC 6000

CACAGACCCC AACCCACAAG AAATAGTATT GGCAAATGTG ACAGAAGATT TTAACATGTG 6060

GAAAAATGAA ATGGTAGAAC AGATGCATAC TGATATAATC AGTTTATGGG ATGAAAGCCT 6120

AAAACCATGT GTAAAATTAA CCCCACTCTG TGTTACTTTA AATTGCACTG ATTTGAAGAA 6180

TGAAACTAAG ACCAATAGTA GTGATGCCAA TAGTAATAGC GGGGAAATAA TGGGGAACGA 6240

AGAGATAAAA AATTGCTCTT TCAATGTCAG CACAGGCGCA CCAGGTAAGG TGCAGAAAGA 6300

ATATGCACTT TTTTATGCAC TTGATATAGT ATCAATAAAG AATGAAAATA ATAGTACCAG 6360

CCATATGTTG ACAAGTTGTA ACACCTCAGT CAGTACACAG GCCTGTCCAA AGGTATCCTT 6420

TGAGCCAATT CCCATACATT ATTGTGCCCC GGCTGGTTTT GCAATTCTAA AATGTAATGA 6480

TAAGAAGTTC AATGGAACAG GACCATGTAA CAATGTCAGC ACAGTACAAT GTACACATGG 6540

AATTAGACCA GTAGTGTCAA CTCAACTGCT GTTAAATGGC AGTGTAGCAG AAGAAGAGGT 6600

AGTACTTAGA TCTGCCAATT TCTCAGACAA TGCTAAAACC ATAATAGTAC AGCTGAACCA 6660

CTCTGTAGAA ATTACTTGTA CAAGACCCAA CTACAATGAA ACAAAGAAAA TCCGTATCCA 6720

CAGAGGATAT GGAAGATCAT TTGTTACAGT AAGAAAATTG GGAGATAGGA AACAAGCACA 6780

TTGTACCATG AATAGAACGA AATGGGACAA CGCTTTAAAA CAGATAGCTA GCAAATTAAG 6840

AGAACAATTT AATAATAAAA CAGCAATAAT CTTTAACCGG TCCTCAGGAG GGGACCTAGA 6900

AATTGAAATG CACAGTTTTA ATTGCGGAGG GGAATTGTTC TACTGTAATA CAACAAAACT 6960

GTTTAATAGT ACTTGGAATG AGACTACAGA GTCAAATGGC AAGGGAGAAA ATATCACACT 7020

CCCATGCAGA ATAAGACAAT TTGTAAACAT GTGGCAGAAA GTAGGAAAAG CAATGTATGC 7080

CCCTCCCAGC GATGGACAAA TTAGGTGTAC ATCAAATATT ACTGGGCTAC TATTAACAAG 7140

AGATGGGGGT CATAATGATA ACAACACTAA CAACGAGACC TTCAGACCGG GAAGAGGAGA 7200

TATGAGGGAC AATTGGAGAA GTGAATTATA TAAATATAAA GTAATAAAAA TTGAACCATT 7260

AGGAGTAGCA CCCACCAAGG CAAAGAGAAG AGTGGTGCAG AGAGAAAAAA GAGCAGTGGG 7320

AATGGTAGGA GCTATGTTCC TTGGGTTCTT GGGAGCAGCA GGAAGCACTA TGGGCGCAGC 7380

GTCATTGACG CTGACGGTAC AGGCCAGACA ATTATTGTCT GGTATAGTGC AGCAGCAGAA 7440

CAATCTGCTG AGAGCTATTG AGGCGCAACA ACATCTGTTG CAACTCACAG TCTGGGGCAT 7500

CAAGCAGCTC CAGGCAAGAG TCCTGGCTGT AGAAAGATAC CTAAAGGATC AACAGCTCCT 7560

GGGGATCTGG GGTTGCTCTG GAAAACTCAT TTGCACCACT ACTGTGCCTT GGAATGCTAG 7620

TTGGAGTAAT AAATCTTTGG ATCAGATTTG GAATAACATG ACCTGGATGG AGTGGGACAG 7680

AGAAATTGCC AATTACACAA ACTTAATACA TCACTTAATT GAAGAATCGC AAAACCAGCA 7740

AGAAAAGAAT GAACAAGAAT TATTGGAATT AGATAAATGG GCAAGTTTGT GGAGTTGGTT 7800

TGACATATCA AACTGGCTGT GGTATATAAA AATATTCATA ATGATAGTAG CAGGCTTAGT 7860

AGGTTTAAGA ATAGTTTTTG CTGTGCTTTC TATAGTAAAT AGAGTTAGGC AGGGATACTC 7920

ACCATTGTCA TTCCAGACCC ACTTCCCAGC TCCGAGGGGA CCCGACAGGC CAGACGGAAT 7980

CGAAGGAGAA GGTGGAGAGA GAGACAGAGA CAGATCCGTG CGATTAGTGG ATGGATTCTT 8040

AGCACTTCTC TGGGAAGACC TGCGCAACCT GTGCCTCTTC AGCTACCACC GCTTGAGAGA 8100

CTTACTCTTG ATTGTAACGA GGATTGTGGA ACTTCTCGGA CGCAGGGGGT GGGAAGCCCT 8160

CAAATATTTG TGGAGTCTCC TACAGTATTG GAGTCAGGAG CTAAAGAATA GTGCTGTCAA 8220

CTTGTTCAAT ACCACAGCTA TAGTAGTAGC TGAGGGGACA GATAGGATCA TAGAAGTAGT 8280

ACAAAGACTT TGTAGAGCTA TTCTCCACAT ACCTAGAAGA ATTAGACAGG GCTTGGAAAG 8340

GATTTTGCTA TAAGATGGGT GGCAAGTGGT CAAAAAGTAG TATAGTTGGA TGGCCTACTA 8400

TAAGGGAAAG AATGAAACGA GCTGGACCAG CAGCAGATGG GGTGGGAGCA GCATCTCGAG 8460

ACCTAGAAAA ACATGGAGCA ATCACAAGTA GCAATACAGC AGCTACCAAT GCTGATTGTG 8520

CCTGGCTAGA AGCACAAGAG GAGGAAGAGG TGGGTTTTCC AGTCAGACCT CAGGTACCTT 8580

TAAGACCAAT GACTTACAAG GCAGGTATAG ATCTTAGCCA CTTTTTAAAA GAAAAGGGGG 8640

GACTGGAAGG GCTAGTTTGG TCCCAAAGAA GACAAGATAT CCTTGATCTG TGGATCTACC 8700

ACACACAAGG CTACTTCCCT GATTGGCAGA ACTACACACC AGGGCCAGGG ATCAGATATC 8760

CACTGACCTT TGGATGGTGC TTCAAGCTAG TACCAGTTGA GCCAGATAAG GTAGAAGAGG 8820

CCAATGAAGG AGAGGACAAC ATCTTGTTAC ACCCTATGTG TCTACATGGA ATGGAGGACG 8880

CGGAGAAAGA AGTGTTAGTG TGGAGATTTG ACAGTAAACT AGCCTTCCAT CACGTAGCCC 8940

GAGAGCTGCA TCCGGAGTAC TACAAAGACT GCTGACATCG AGCTTTCTAC AAGGGACTTT 9000

CCGCTGGGGA CTTTCCAGGG AGGCGTGATC TGGGTGGGAC TGGGGAGTGG CGTGCCCTCA 9060

GATGCTGCAT ATAAGCAGCT GCTTTTGCCT GTACTGGGTC TCTCTGGTTA GACCAGATCT 9120

GAGCCTGGGA GCTCTCTGGC TAGCTAGGGA ACCCACTGCT TAAGCCTCAA TAAAGCTTGC 9180

CTTGAGTGCT TCA 9193

(SEQ ID NO:12)

HIVNCFULL7.SEQ

ggtctctctggttagaccagatctgagcctgggagctctctggctagctagggaacccactgcttaagcc
tcaataaagcttgccttgagtgcttcaagtagtgtgtgcccgtctgttgtgtgactctggtaactagaga
tccctcagacccttaaagtcagtgtggaaaaatctctagcagtggcgcccgaacagggacgcgaaagcga
aagtagaaccagagaagctctctcgacgcaggactcggcttgctgaagcgcgcacagcaagaggcgaggg
gaggcgactggtgagtacgccaatatttttgactagcggaggctagaaggagagagatgggtgcgagagc
gtcggtattaagcgggggagatttagatcgatgggaaaaaattcggttaaggccaggggggaaagaaaaaa
tatatgttaaaacatatagtatgggcaagcagggagctagaacgattcgcagttaatcctggcctgttag
aaacagctgaaggctgtagacaaatactagaacagctacagccatcccttcagacaggatcagaagaact
tagatccttatataatacaatagcaacccctctattgtgtgcatcaaaaaatagaggtaaaagacaccaag
gaagctttagaaaaaatagaggaagagcaaaacaaaagtaagaaaaaagtacagcaagcagcagctgcaa
ctggcacaggaaacagcagccaggtcagccaaaattaccctatagtgcagaacctacaggggcaaatggt
acatcaggccatatcacctagaactttaaatgcatgggtaaaagtaatagaagagaaggctttcagccca
gaagtgataccatgttttcagcattatcagaaggagccaccccacaagatttaaacaccatgctaaaca
cagtgggggacatcaagcagccatgcaaatgttaaaagagaccatcaatgaggaagctgcagaatggga
tagagtgcatccagtgcaggcagggcttattgcaccaggccagatgagagaaccaaggggaagtgacata
gcaggaactactagtacccttcaggaacaaataggatggatgacaagtaatccacctatcccagtaggag
aaatttataaaagatggataatttgggcttaaataaaatagtaagaatgtatagccctatcagcattct
ggatataagacaaggaccaaaagaacccttagagactatgtagaccggttctataaaactctaagagcc
gagcaagcttcacaggaagtaaaaaattggatgacagaaaccttgttggtccaaaatgcgaacccagatt
gtaagactattttaaaagcattaggaccaggagctacattagaagaaatgatgacagcatgtcagggagt
gggaggacccggccataaggcaagggttttagctgaagcaatgagccaagtaacaaattcagctgccata
atgatgcagagaggcaattttaaaaaccaaagaaagatggttaagtgtttcaattgtggcaaagaagggc
atgtagccagaaattgcagggcccctagaaaaaagggctgttggaaatgtggaaaggaaggacaccaaat
gaaagattgtattgaaagacaggctaatttttttagggaaaatctggccttcccacaagggaaggccaggg
aattttcttcagagcagaccagagccaacagccccaccagaagagagcctcaggtctgggatagagacaa
caactccctctcagaagcaggagccaatagacaaggaagtgtatcctttaacttccctcagatcactctt
tggcaacgaccctcgtcacaataaagatagggggcaactaaaggaagctctattagatacaggagcag
atgatacagtattagaagaaatgaatttgccaggaagatggaaaccaaaaatgatagggggaattggagg
ttttatcaaagtaagacagtatgatcagatacccatagaaatctgtggacataaagctataggtacagta
ttagtaggacctacacctgtcaacataattggaagaaatctgttgactcagattggttgcactttaaatt
ttcccattagtcctattgaaactgtaccagtaaaattaaagccaggaatggatggcccaaaagttaaaca
atggccattgacagaagaaaaaataaaagcattaatagaaatttgtacagaaatggaaaaggaagggaaa
atttcaaaaattgggcctgaaaatccatacaatactccagtatttgccataaagaaaaaagacagtacta
aatggagaaaattagtagatttcagagaacttaataagagaactcaagacttctgggaagttcaattagg
aataccacatcccgcagggttaaaaaagaaaaaatcagtaacagtactagatgtgggtgatgcatatttt
tcagttcccttagatgaagatttcaggaagtatactgcatttaccatacctagtataaacaatgagacac
cagggattagatatcagtacaatgtgctgccacagggatggaaaggatcaccagcaatattccaaagtag
catgacaaaaatcttagagccttttagaaaacagaatccagacatagttatctatcaatatatggatgat
ttgtatgtaggatctgacttagaaatagagcagcatagaacaaaaatagaggaactgagacagcatctgt
tgaagtggggatttaccacaccagacaaaaaacatcagaaggaacctccattcctctggatgggttatga
actccatcctgataaatggacagtacagcctatagtgctgccagagaaagacagctggactgtcaatgac
atacagaagttagtgggaaaattaaattgggcaagtcagatttacccagggattaaagtaaggcaattat
gtaaactccttaggggagccaaagcactaacagatgtaataccactaacagaaggagcagagctagaact
ggcagaaaacagagagattctaaaagaaccagtacatggagtgtattatgacccatcaaaagacttaata
gcagaaatacagaagcagggacaaggtcaatggacatatcagatttatcaagagccatttaaaaatctga
gaacaggaaaatatgcaagaacgaggggtgcccacactaatgatataaaacaattaacagaggcagtgca
aaaaatagccacagaaagcatagtaatatggggaaagactcctaaatttaaactgcccatacaaaaggaa
acatgggaaacatggtggacagagtattggcaagctacctggattcctgagtgggagtttgtcaataccc
ctcctttagtaaaattatggtaccagttagagaaagaacccatagtaggagcagaaactttctatgtaga
tggggcagctaacagggagactaaattaggaaaagcaggatatgttactaacagaggaagacaaaaggta
gtccccctgactaacacaacaaatcagaagactgagttacaagcaattcatctagctttacaggattcag
gatcagaagtaaacatagtaacagactcacaatatgcattaggaatcattcaagcacaaccagatcaaag
tgaatcagagttagtcaatcaaataatagaacagttaataaaaaaggaaaaggtctatctggcatgggta
ccagcacacaaaggaattggaggaaatgaacaagtagataaattagtcagtgctggaatcaggaaagtac
tatttctagatggaatagataaggcccaagaagaacatgagaaatatcataataattggagagcaatggc
tagtgattttaacctgccacctgtagtagcaaaggaaatagtagccagctgtgataaatgtcagctaaaa

ggagaagccatgcatggacaagtagactgtagtccaggaatatggcaactagattgtacacatttagaag
gaaaagttatcctagtagcagttcatgtagccagtggatatatagaagcagaagttattccagcagaaac
agggcaggaaacagcatactttctcttaaaattagcaggaagatggccagtaaaaacaatacacacagat
aatggcagcaatttcaccagtgctgcggttaaggccgcctgttggtgggcgggagtcaagcaggaatttg
gaattccctacaatccccaaagtcaaggagtagtagaatctatgaataaagaattaaagaaaattatagg
acaggtaagagatcaagctgaacatcttaagacagcagtacaaatggcagtatttgtccacaattttaaa
agaaaaggggggattgggggtacagtgcaggggaaagaataatagacataatagcaacagacatacaaa
ctagagaactacaaaaacaaattacaaaaattcaaaatttcgggtttattacagggacagcagagaccc
actttggaaaggaccagcaaagctcctctggaaaggtgaaggggcagtagtaatacaagataatagtgac
ataaaagtagtgccaagaagaaaagcaaagatcattagggattatggaaaacagatggcaggtgatgatt
gtgtggcaagtagacaggatgaggattagaacatggacaagtttagtaaaacatcatgtatatgtttcaa
agaaagctaggggatggttttatagacatcactatgaaagcactcatccaaaaataagttcagaaataca
cattccactaggggatgctagattggtagtaacaacatattggggtctgcatacaggagaaagagaatgg
catttgggtcatggagtttccatagaatggaggaaaaggagctatagcacacaagtagaccctgaactag
cagaccaactaattcatctgtattactttgattgttttgcagactctgctataagaaaggccttattagg
acacatagttagccctaggtgtgcatatcaagcaggacattccaaggtaggatctctacaatacttggca
ctatcagccttaataacaccaaagaagataaagccacctttgcctagtgttacaaaactgacagaggata
gatggaacaagccccagaagaccaagggccacagagggagccacacaatgaatggacactagagctttta
gaggagcttaagacagaagctgttagacattttcctaggccatggctccatagcttaggacaacatatct
atgaaacttatggagatacttgggcaggagtggaagccataataagaattctgcaacaattgctgtttat
tcatttcagaattgggtgtcaacatagcagaataggcattattcaacagaggagagcaagaagaaatgga
gccagtagatcctaatctagagccctggaagcatccaggaagtcagcctaaaactgcttgtaccaaatgc
tattgtaaaaagtgttgctttcattgccaagtttgtttcacaacaaaagccttaggcatctcctatggca
ggaagaagcggagacagcgacgaagagctcctcagggcagccaggctcatcaagcttctctatcagagca
gtaagtagtatatgtaatgcaacttatattaatagtaacaatagtagctttagtagtaacattaataata
gcaatagttgtgtggtccatagtactcatagaatataggaaaatattaagacaaaggaaaatagacaagt
taattaatagactagtagaaagagcagaagacagtggcaatgagagtgaaggagaactatcagcacttgt
ggagatgggggtggaaatggggcatcatgctccttgggatgttaatgatctgtagtgctacagaaaaatt
gtgggtcacagtctattatggggtacctgtgtggaaggaagcaaccaccactctattttgtgcatcagat
gctaaagcatatgaagaagaggtacataatgtttgggccacacatgcctgtgtacccacagaccccaacc
cacaagaaatagtattggcaaatgtgacagaagattttaacatgtggaaaaatgaaatggtagaacagat
gcatactgatataatcagtttatgggatgaaagcctaaaaccatgtgtaaaattaaccccactctgtgtt
actttaaattgcactgatttgaagaatgaaactaagaccaatagtagtgatgccaatagtaatagcgggg
aaataatggggaacgaagagataaaaaattgctctttcaatgtcagcacaggcgcaccaggtaaggtgca
gaaagaatatgcacttttttatgcacttgatatagtatcaataaagaatgaaaataatagtaccagccat
atgttgacaagttgtaacacctcagtcagtacacaggcctgtccaaaggtatcctttgagccaattccca
tacattattgtgccccggctggttttgcaattctaaaatgtaatgataagaagttcaatggaacaggacc
atgtaacaatgtcagcacagtacaatgtacacatggaattaggccagtagtatcaactcaactgctgtta
aatggcagtctagcagaagaagaggtagtacttagatctgccaatttctcagacaatgctaaaaccataa
tagtacagctgaaccactctgtagaaattacttgtacaagacccaactacaatgaaacaaaaagaatccg
tatccacagaggatatggaagatcatttgttacagtaagaaaattgggagataggaaacaagcacattgt
accatgaatagaacgaaatgggacaacgctttaaaacagatagctagcaaattaagagaacaatttaata
aaacagcaataatctttaaccggtcctcaggaggggacctagaaattgaaatgcacagttttaattgcgg
aggggaattgttctactgtaatacaacaaaactgtttaatagtacttggaatgagactacagagtcaaat
ggcaagggagaaaatatcacactcccatgcagaataagacaatttgtaaacatgtggcagaaagtaggaa
aagcaatgtatgccccctcccagcgatggacaaattaggtgtacatcaaatattactgggctactattaac
aagagatgggggtggtcctagtgataacaaaaccgacaagaccttcagaccaggaggaggagatatgagg
gacaattggagaagtgaattatataaatataaagtaataaaaattgaaccattaggagtagcacccacca
aggcaaagagaagagtggtgcaaagagaaaaaagagcagtgggaatggtaggagctatgttccttgggtt
cttgggagcagcaggaagcactatgggcgcagcgtcattgacgctgacggtacaggccagacaattattg
tctggtatagtgcagcagcagaacaatctgctgagagctattgaggcgcaacaacatctgttgcaactca
cagtctggggcatcaagcagctccaggcaagagtcctggctgtagaaagatacctaaaggatcaacagct
cctggggatctggggttgctctggaaaactcatttgcaccactactgtgccttggaatgctagttggagt
aataaatctctggatcagatttggaataacatgacctggttggagtgggacagagaaattgccaattaca
caaacttaatacatcacttaattgaagaatcgcaaaaccagcaagaaaagaatgaacaagaattattgga
attagataaatgggcaagtttgtggagttggtttgacatatcaaactggctgtggtatataaaaatattc
ataatgatagtagcaggcttagtaggtttaagaatagtttttgctgtgctttctatagtaaatagagtta
ggcagggatactcaccattgtcattccagacccacttcccagctccgaggggacccgacaggccagacgg

aatcgaaggagaaggtggagagagagacagagacagatccgtgcgattagtggatggattcttagcactt
ctctgggaagacctgcgcaacctgtgcctcttcagctaccaccgcttgagagacttactcttgattgtaa
cgaggattgtggaacttctgggacgcaggggtgggaagccctcaaatattggtggagtctcctacagta
ttggagtcaggagctaaagaatagtgctgtcaacttgttcaataccagagctatagtagtagctgagggg
acagataggatcatagaagtagtacaaagactttgtagagctattctccacatacctagaagaattagac
agggcttggaaagatttttgctataagatgggtggcaagtggtcaaaaagtagtatagttggatggccta
ctataagggaaagaatgaaacgagctggaccagcagcagatggggtgggagcagcatctcgagacctaga
aaaacatcgagcaatcacaagtagcaatacagcagctaccaatgctgattgtgcctggctagaagcacaa
gaggaggaagaggtgggttttccagtcagacctcaggtacctttaagaccaatgacttacaaggcaggta
tagatcttagccacttttaaaagaaaaggggggactggaagggctagtttggtcccaaagaagacaaga
tatccttgatctgtggatctaccacacacaaggctacttccctgattggcagaactacacaccagggcca
gggatcagatatccactgacctttggatggtgcttcaagctagtaccagttgagccagataaggtagaag
aggccaatgaaggagaggacaacatcttgttacaccctatgtgtctacatggaatggaggacgcggagaa
agaagtgttagtgtggagatttgacagtaaactagccttccatcacgtagcccgagagctgcatccggag
tactacaaagactgctgacatcgagctttctacaagggactttccgctggggactttccagggaggcgtg
atctgggtgggactggggagtggcgtgccctcagatgctgcatataagcagctgcttttgcctgtactgg
gtctctctggttagaccagatctgagcctgggagctctctggctagctagggaacccactgcttaagcct
caataaagcttgccttgagtgcttca

FOR SEQ ID NO:13 THROUGH SEQ ID NO:17, PLEASE SEE TABLE 5.

(2) INFORMATION FOR SEQ ID NO:18

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "Oligonucleotide."

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18

AAATCTCTAG CAGTGGCGCC CGAACAG 27

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "Oligonucleotide."

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

GCACTCAAGG CAAGCTTTAT TGAGGCT 27

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 31 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "Oligonucleotide."

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

CACACACAAG GCTACTTCCC TGATTGGCAG A 31

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 33 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "Oligonucleotide."

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

ATGGAACAAG CCCCAGAAGA CCAAGGGCCA CAG 33

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 31 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "Oligonucleotide."

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

GGTCTGAGGG ATCTCTAGTT ACCAGAGTCA C 31

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2404 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

Met Glu Thr Ala Arg Gly Val Ala Leu Leu Tyr Ser Gly Leu Ala Ser
1 5 10 15

Asn Thr Tyr Arg Gly Leu Asn His Ile Ser Leu Glu Thr Arg Pro Ala
20 25 30

Arg Gly Thr Arg Pro Gly Leu Tyr Thr Arg Pro Leu Tyr Ser Thr Arg
35 40 45

Pro Gly Leu Tyr Ile Leu Glu Met Glu Thr Leu Glu Leu Glu Gly Leu
50 55 60

Tyr Met Glu Thr Leu Glu Met Glu Thr Ile Leu Glu Cys Tyr Ser Ser
65 70 75 80

Glu Arg Ala Leu Ala Thr His Arg Gly Leu Leu Tyr Ser Leu Glu Thr
85 90 95

Arg Pro Val Ala Leu Thr His Arg Val Ala Leu Thr Tyr Arg Thr Tyr
100 105 110

Arg Gly Leu Tyr Val Ala Leu Pro Arg Val Ala Leu Thr Arg Pro Leu
115 120 125

Tyr Ser Gly Leu Ala Leu Ala Thr His Arg Thr His Arg Thr His Arg
130 135 140

Leu Glu Pro His Glu Cys Tyr Ser Ala Leu Ala Ser Glu Arg Ala Ser
145 150 155 160

Pro Ala Leu Ala Leu Tyr Ser Ala Leu Ala Thr Tyr Arg Gly Leu Gly
165 170 175

Leu Gly Leu Val Ala Leu His Ile Ser Ala Ser Asn Val Ala Leu Thr
180 185 190

Arg Pro Ala Leu Ala Thr His Arg His Ile Ser Ala Leu Ala Cys Tyr
195 200 205

Ser Val Ala Leu Pro Arg Thr His Arg Ala Ser Pro Pro Arg Ala Ser
210 215 220

Asn Pro Arg Gly Leu Asn Gly Leu Ile Leu Glu Val Ala Leu Leu Glu
225 230 235 240

Ala Leu Ala Ala Ser Asn Val Ala Leu Thr His Arg Gly Leu Ala Ser
245 250 255

Pro Pro His Glu Ala Ser Asn Met Glu Thr Thr Arg Pro Leu Tyr Ser
260 265 270

Ala Ser Asn Gly Leu Met Glu Thr Val Ala Leu Gly Leu Gly Leu Asn
275 280 285

Met Glu Thr His Ile Ser Thr His Arg Ala Ser Pro Ile Leu Glu Ile
290 295 300

Leu Glu Ser Glu Arg Leu Glu Thr Arg Pro Ala Ser Pro Gly Leu Ser
305 310 315 320

Glu Arg Leu Glu Leu Tyr Ser Pro Arg Cys Tyr Ser Val Ala Leu Leu
325 330 335

Tyr Ser Leu Glu Thr His Arg Pro Arg Leu Glu Cys Tyr Ser Val Ala
340 345 350

Leu Thr His Arg Leu Glu Ala Ser Asn Cys Tyr Ser Thr His Arg Ala
355 360 365

Ser Pro Leu Glu Leu Tyr Ser Ala Ser Asn Gly Leu Thr His Arg Leu
370 375 380

Tyr Ser Thr His Arg Ala Ser Asn Ser Glu Arg Ser Glu Arg Ala Ser
385 390 395 400

Pro Ala Leu Ala Ala Ser Asn Ser Glu Arg Ala Ser Asn Ser Glu Arg
405 410 415

Gly Leu Tyr Gly Leu Ile Leu Glu Met Glu Thr Gly Leu Tyr Ala Ser
420 425 430

Asn Gly Leu Gly Leu Ile Leu Glu Leu Tyr Ser Ala Ser Asn Cys Tyr
435 440 445

Ser Ser Glu Arg Pro His Glu Ala Ser Asn Val Ala Leu Ser Glu Arg
450 455 460

Thr His Arg Gly Leu Tyr Ala Leu Ala Pro Arg Gly Leu Tyr Leu Tyr
465 470 475 480

Ser Val Ala Leu Gly Leu Asn Leu Tyr Ser Gly Leu Thr Tyr Arg Ser
485 490 495

Glu Arg Leu Glu Pro His Glu Thr Tyr Arg Ala Leu Ala Leu Glu Ala
500 505 510

Ser Pro Ile Leu Glu Val Ala Leu Ser Glu Arg Ile Leu Glu Leu Tyr
515 520 525

Ser Ala Ser Asn Gly Leu Ala Ser Asn Ala Ser Asn Ser Glu Arg Thr
530 535 540

His Arg Ser Glu Arg His Ile Ser Met Glu Thr Leu Glu Thr His Arg
545 550 555 560

Ser Glu Arg Cys Tyr Ser Ala Ser Asn Thr His Arg Ser Glu Arg Val
565 570 575

Ala Leu Ser Glu Arg Thr His Arg Gly Leu Asn Ala Leu Ala Cys Tyr
580 585 590

Ser Pro Arg Leu Tyr Ser Val Ala Leu Ser Glu Arg Pro His Glu Gly
595 600 605

Leu Pro Arg Ile Leu Glu Pro Arg Ile Leu Glu His Ile Ser Thr Tyr
610 615 620

Arg Cys Tyr Ser Ala Leu Ala Pro Arg Ala Leu Ala Gly Leu Tyr Pro
625 630 635 640

His Glu Ala Leu Ala Ile Leu Glu Leu Glu Leu Tyr Ser Cys Tyr Ser
645 650 655

Ala Ser Asn Ala Ser Pro Leu Tyr Ser Leu Tyr Ser Pro His Glu Ala
660 665 670

Ser Asn Gly Leu Tyr Thr His Arg Gly Leu Tyr Pro Arg Cys Tyr Ser
675 680 685

Ala Ser Asn Ala Ser Asn Val Ala Leu Ser Glu Arg Thr His Arg Val
690 695 700

Ala Leu Gly Leu Asn Cys Tyr Ser Thr His Arg His Ile Ser Gly Leu
705 710 715 720

Tyr Ile Leu Glu Ala Arg Gly Pro Arg Val Ala Leu Val Ala Leu Ser
725 730 735

Glu Arg Thr His Arg Gly Leu Asn Leu Glu Leu Glu Leu Glu Ala Ser
740 745 750

Asn Gly Leu Tyr Ser Glu Arg Leu Glu Ala Leu Ala Gly Leu Gly Leu
755 760 765

Gly Leu Val Ala Leu Val Ala Leu Leu Glu Ala Arg Gly Ser Glu Arg
770 775 780

Ala Leu Ala Ala Ser Asn Pro His Glu Ser Glu Arg Ala Ser Pro Ala
785 790 795 800

Ser Asn Ala Leu Ala Leu Tyr Ser Thr His Arg Ile Leu Glu Ile Leu
805 810 815

Glu Val Ala Leu Gly Leu Asn Leu Glu Ala Ser Asn His Ile Ser Ser
820 825 830

Glu Arg Val Ala Leu Gly Leu Ile Leu Glu Thr His Arg Cys Tyr Ser
835 840 845

Thr His Arg Ala Arg Gly Pro Arg Ala Ser Asn Thr Tyr Arg Ala Ser
850 855 860

Asn Gly Leu Thr His Arg Leu Tyr Ser Leu Tyr Ser Ile Leu Glu Ala
865 870 875 880

Arg Gly Ile Leu Glu His Ile Ser Ala Arg Gly Gly Leu Tyr Thr Tyr
885 890 895

Arg Gly Leu Tyr Ala Arg Gly Ser Glu Arg Pro His Glu Val Ala Leu
900 905 910

Thr His Arg Val Ala Leu Ala Arg Gly Leu Tyr Ser Leu Glu Gly Leu
915 920 925

Tyr Ala Ser Pro Ala Arg Gly Leu Tyr Ser Gly Leu Asn Ala Leu Ala
930 935 940

His Ile Ser Cys Tyr Ser Thr His Arg Met Glu Thr Ala Ser Asn Ala
945 950 955 960

Arg Gly Thr His Arg Leu Tyr Ser Thr Arg Pro Ala Ser Pro Ala Ser
965 970 975

Asn Ala Leu Ala Leu Glu Leu Tyr Ser Gly Leu Asn Ile Leu Glu Ala
980 985 990

Leu Ala Ser Glu Arg Leu Tyr Ser Leu Glu Ala Arg Gly Gly Leu Gly
995 1000 1005

Leu Asn Pro His Glu Ala Ser Asn Leu Tyr Ser Thr His Arg Ala Leu
1010 1015 1020

Ala Ile Leu Glu Ile Leu Glu Pro His Glu Ala Ser Asn Ala Arg Gly
1025 1030 1035 1040

Ser Glu Arg Ser Glu Arg Gly Leu Tyr Gly Leu Tyr Ala Ser Pro Leu
1045 1050 1055

Glu Gly Leu Ile Leu Glu Gly Leu Met Glu Thr His Ile Ser Ser Glu
1060 1065 1070

Arg Pro His Glu Ala Ser Asn Cys Tyr Ser Gly Leu Tyr Gly Leu Tyr
1075 1080 1085

Gly Leu Leu Glu Pro His Glu Thr Tyr Arg Cys Tyr Ser Ala Ser Asn
1090 1095 1100

Thr His Arg Thr His Arg Leu Tyr Ser Leu Glu Pro His Glu Ala Ser
1105 1110 1115 1120

Asn Ser Glu Arg Thr His Arg Thr Arg Pro Ala Ser Asn Gly Leu Thr
1125 1130 1135

His Arg Thr His Arg Gly Leu Ser Glu Arg Ala Ser Asn Gly Leu Tyr
1140 1145 1150

Leu Tyr Ser Gly Leu Tyr Gly Leu Ala Ser Asn Ile Leu Glu Thr His
1155 1160 1165

Arg Leu Glu Pro Arg Cys Tyr Ser Ala Arg Gly Ile Leu Glu Ala Arg
1170 1175 1180

Gly Gly Leu Asn Pro His Glu Val Ala Leu Ala Ser Asn Met Glu Thr
1185 1190 1195 1200

Thr Arg Pro Gly Leu Asn Leu Tyr Ser Val Ala Leu Gly Leu Tyr Leu
1205 1210 1215

Tyr Ser Ala Leu Ala Met Glu Thr Thr Tyr Arg Ala Leu Ala Pro Arg
1220 1225 1230

Pro Arg Ser Glu Arg Ala Ser Pro Gly Leu Tyr Gly Leu Asn Ile Leu
1235 1240 1245

Glu Ala Arg Gly Cys Tyr Ser Thr His Arg Ser Glu Arg Ala Ser Asn
1250 1255 1260

Ile Leu Glu Thr His Arg Gly Leu Tyr Leu Glu Leu Glu Leu Glu Thr
1265 1270 1275 1280

His Arg Ala Arg Gly Ala Ser Pro Gly Leu Tyr Gly Leu Tyr Gly Leu
1285 1290 1295

Tyr Pro Arg Ser Glu Arg Ala Ser Pro Ala Ser Asn Leu Tyr Ser Ala
1300 1305 1310

Ser Pro Leu Tyr Ser Gly Leu Thr His Arg Pro His Glu Ala Arg Gly
1315 1320 1325

Pro Arg Gly Leu Tyr Gly Leu Tyr Gly Leu Tyr Ala Ser Pro Met Glu
1330 1335 1340

Thr Ala Arg Gly Ala Ser Pro Ala Ser Asn Thr Arg Pro Ala Arg Gly
1345 1350 1355 1360

Ser Glu Arg Gly Leu Leu Glu Thr Tyr Arg Leu Tyr Ser Thr Tyr Arg
1365 1370 1375

Leu Tyr Ser Val Ala Leu Ile Leu Glu Leu Tyr Ser Ile Leu Glu Gly
1380 1385 1390

Leu Pro Arg Leu Glu Gly Leu Tyr Val Ala Leu Ala Leu Ala Pro Arg
1395 1400 1405

Thr His Arg Leu Tyr Ser Ala Leu Ala Leu Tyr Ser Ala Arg Gly Ala
1410 1415 1420

Arg Gly Val Ala Leu Val Ala Leu Gly Leu Asn Ala Arg Gly Gly Leu
1425 1430 1435 1440

Leu Tyr Ser Ala Arg Gly Ala Leu Ala Val Ala Leu Gly Leu Tyr Met
1445 1450 1455

Glu Thr Val Ala Leu Gly Leu Tyr Ala Leu Ala Met Glu Thr Pro His
1460 1465 1470

Glu Leu Glu Gly Leu Tyr Pro His Glu Leu Glu Gly Leu Tyr Ala Leu
1475 1480 1485

Ala Ala Leu Ala Gly Leu Tyr Ser Glu Arg Thr His Arg Met Glu Thr
1490 1495 1500

Gly Leu Tyr Ala Leu Ala Ala Leu Ala Ser Glu Arg Leu Glu Thr His
1505 1510 1515 1520

Arg Leu Glu Thr His Arg Val Ala Leu Gly Leu Asn Ala Leu Ala Ala
1525 1530 1535

Arg Gly Gly Leu Asn Leu Glu Leu Glu Ser Glu Arg Gly Leu Tyr Ile
1540 1545 1550

Leu Glu Val Ala Leu Gly Leu Asn Gly Leu Asn Gly Leu Asn Ala Ser
1555 1560 1565

Asn Ala Ser Asn Leu Glu Leu Glu Ala Arg Gly Ala Leu Ala Ile Leu
1570 1575 1580

Glu Gly Leu Ala Leu Ala Gly Leu Asn Gly Leu Asn His Ile Ser Leu
1585 1590 1595 1600

Glu Leu Glu Gly Leu Asn Leu Glu Thr His Arg Val Ala Leu Thr Arg
1605 1610 1615

Pro Gly Leu Tyr Ile Leu Glu Leu Tyr Ser Gly Leu Asn Leu Glu Gly
1620 1625 1630

Leu Asn Ala Leu Ala Ala Arg Gly Val Ala Leu Leu Glu Ala Leu Ala
1635 1640 1645

Val Ala Leu Gly Leu Ala Arg Gly Thr Tyr Arg Leu Glu Leu Tyr Ser
1650 1655 1660

Ala Ser Pro Gly Leu Asn Gly Leu Asn Leu Glu Leu Glu Gly Leu Tyr
1665 1670 1675 1680

Ile Leu Glu Thr Arg Pro Gly Leu Tyr Cys Tyr Ser Ser Glu Arg Gly
1685 1690 1695

Leu Tyr Leu Tyr Ser Leu Glu Ile Leu Glu Cys Tyr Ser Thr His Arg
1700 1705 1710

Thr His Arg Thr His Arg Val Ala Leu Pro Arg Thr Arg Pro Ala Ser
1715 1720 1725

Asn Ala Leu Ala Ser Glu Arg Thr Arg Pro Ser Glu Arg Ala Ser Asn
1730 1735 1740

Leu Tyr Ser Ser Glu Arg Leu Glu Ala Ser Pro Gly Leu Asn Ile Leu
1745 1750 1755 1760

Glu Thr Arg Pro Ala Ser Asn Ala Ser Asn Met Glu Thr Thr His Arg
1765 1770 1775

Thr Arg Pro Leu Glu Gly Leu Thr Arg Pro Ala Ser Pro Ala Arg Gly
1780 1785 1790

Gly Leu Ile Leu Glu Ala Leu Ala Ala Ser Asn Thr Tyr Arg Thr His
1795 1800 1805

Arg Ala Ser Asn Leu Glu Ile Leu Glu His Ile Ser His Ile Ser Leu
1810 1815 1820

Glu Ile Leu Glu Gly Leu Gly Leu Ser Glu Arg Gly Leu Asn Ala Ser
1825 1830 1835 1840

Asn Gly Leu Asn Gly Leu Asn Gly Leu Leu Tyr Ser Ala Ser Asn Gly
1845 1850 1855

Leu Gly Leu Asn Gly Leu Leu Glu Leu Glu Gly Leu Leu Glu Ala Ser
1860 1865 1870

Pro Leu Tyr Ser Thr Arg Pro Ala Leu Ala Ser Glu Arg Leu Glu Thr
1875 1880 1885

Arg Pro Ser Glu Arg Thr Arg Pro Pro His Glu Ala Ser Pro Ile Leu
1890 1895 1900

Glu Ser Glu Arg Ala Ser Asn Thr Arg Pro Leu Glu Thr Arg Pro Thr
1905 1910 1915 1920

Tyr Arg Ile Leu Glu Leu Tyr Ser Ile Leu Glu Pro His Glu Ile Leu
1925 1930 1935

Glu Met Glu Thr Ile Leu Glu Val Ala Leu Ala Leu Ala Gly Leu Tyr
1940 1945 1950

Leu Glu Val Ala Leu Gly Leu Tyr Leu Glu Ala Arg Gly Ile Leu Glu
1955 1960 1965

Val Ala Leu Pro His Glu Ala Leu Ala Val Ala Leu Leu Glu Ser Glu
1970 1975 1980

Arg Ile Leu Glu Val Ala Leu Ala Ser Asn Ala Arg Gly Val Ala Leu
1985 1990 1995 2000

Ala Arg Gly Gly Leu Asn Gly Leu Tyr Thr Tyr Arg Ser Glu Arg Pro
2005 2010 2015

Arg Leu Glu Ser Glu Arg Pro His Glu Gly Leu Asn Thr His Arg His
2020 2025 2030

Ile Ser Pro His Glu Pro Arg Ala Leu Ala Pro Arg Ala Arg Gly Gly
2035 2040 2045

Leu Tyr Pro Arg Ala Ser Pro Ala Arg Gly Pro Arg Ala Ser Pro Gly
2050 2055 2060

Leu Tyr Ile Leu Glu Gly Leu Gly Leu Tyr Gly Leu Gly Leu Tyr Gly
2065 2070 2075 2080

Leu Tyr Gly Leu Ala Arg Gly Ala Ser Pro Ala Arg Gly Ala Ser Pro
2085 2090 2095

Ala Arg Gly Ser Glu Arg Val Ala Leu Ala Arg Gly Leu Glu Val Ala
2100 2105 2110

Leu Ala Ser Pro Gly Leu Tyr Pro His Glu Leu Glu Ala Leu Ala Leu
2115 2120 2125

Glu Leu Glu Thr Arg Pro Gly Leu Ala Ser Pro Leu Glu Ala Arg Gly
2130 2135 2140

Ala Ser Asn Leu Glu Cys Tyr Ser Leu Glu Pro His Glu Ser Glu Arg
2145 2150 2155 2160

Thr Tyr Arg His Ile Ser Ala Arg Gly Leu Glu Ala Arg Gly Ala Ser
2165 2170 2175

Pro Leu Glu Leu Glu Leu Glu Ile Leu Glu Val Ala Leu Thr His Arg
2180 2185 2190

Ala Arg Gly Ile Leu Glu Val Ala Leu Gly Leu Leu Glu Leu Glu Gly
2195 2200 2205

Leu Tyr Ala Arg Gly Ala Arg Gly Gly Leu Tyr Thr Arg Pro Gly Leu
2210 2215 2220

Ala Leu Ala Leu Glu Leu Tyr Ser Thr Tyr Arg Thr Arg Pro Thr Arg
2225 2230 2235 2240

Pro Ser Glu Arg Leu Glu Leu Glu Gly Leu Asn Thr Tyr Arg Thr Arg
2245 2250 2255

Pro Ser Glu Arg Gly Leu Asn Gly Leu Leu Glu Leu Tyr Ser Ala Ser
2260 2265 2270

Asn Ser Glu Arg Ala Leu Ala Val Ala Leu Ala Ser Asn Leu Glu Pro
2275 2280 2285

His Glu Ala Ser Asn Thr His Arg Ala Arg Gly Ala Leu Ala Ile Leu
2290 2295 2300

Glu Val Ala Leu Val Ala Leu Ala Leu Ala Gly Leu Gly Leu Tyr Thr
2305 2310 2315 2320

His Arg Ala Ser Pro Ala Arg Gly Ile Leu Glu Ile Leu Glu Gly Leu
2325 2330 2335

Val Ala Leu Val Ala Leu Gly Leu Asn Ala Arg Gly Leu Glu Cys Tyr
2340 2345 2350

Ser Ala Arg Gly Ala Leu Ala Ile Leu Glu Leu Glu His Ile Ser Ile
2355 2360 2365

Leu Glu Pro Arg Ala Arg Gly Ala Arg Gly Ile Leu Glu Ala Arg Gly
2370 2375 2380

Gly Leu Asn Gly Leu Tyr Leu Glu Gly Leu Ala Arg Gly Pro His Glu
2385 2390 2395 2400

Leu Glu Leu Glu

We claim:

1. An isolated human immunodeficiency virus type 1 (HIV-1), which is HIV-1_{JC}, identified by a nucleotide sequence as given in SEQ ID NO:11, or which is HIV-1_{NC}, identified by a nucleotide sequence as given in SEQ ID NO:12.

2. A biological sample comprising the HIV-1_{JC} of claim 1.

3. A biological sample comprising the HIV-1_{NC} of claim 1.

4. The biological sample of claim 1 wherein said sample is a sample of blood.

5. A biologically pure culture of host cells comprising the HIV-1_{JC} or HIV-1_{NC} of claim 1.

6. The culture of claim 5 wherein the host cells are peripheral blood mononuclear cells.

7. A composition comprising an antigenic preparation of the HIV-1_{JC} of claim 1.

8. A composition comprising an antigenic preparation of the HIV-1_{NC} of claim 1.

9. A kit for detecting the presence of HIV-1 antibodies comprising an antigenic preparation of the HIV-1_{JC} or HIV-1_{NC} of claim 1.

10. An immunogenic composition comprising an antigenic preparation of the HIV-1_{JC} or HIV-1_{NC} of claim 1 and a pharmaceutically acceptable carrier.

11. A vaccine comprising an antigenic preparation of the HIV-1_{JC} or HIV-1_{NC} of claim 1.

12. An isolated DNA molecule comprising a nucleotide sequence encoding an infectious molecular clone for HIV-1_{JC} or an antigenic fragment thereof or for HIV-1_{NC} or an antigenic fragment thereof.

13. The isolated DNA molecule of claim 12, wherein the DNA comprises a nucleotide sequence encoding an HIV-1_{JC} envelope protein as given in SEQ ID NO:2, an HIV-1_{JC} gag, nef or p24 protein, or an antigenic fragment of any of the foregoing..

14. The isolated DNA molecule of claim 13, wherein the DNA comprises the nucleotide sequence of SEQ ID NO:1 or a portion thereof specifying an antigenic fragment of said HIV-1_{JC} envelope protein.

15. The isolated DNA molecule of claim 12, wherein the DNA comprises a nucleotide sequence encoding an HIV-1_{NC} envelope protein as given in SEQ ID NO:23, an HIV-1_{NC} gag, nef or p24 protein, or an antigenic fragment of any of the foregoing.

16. A method of inducing antibodies to HIV-1_{JC} in a nonhuman mammalian subject comprising the steps of (a) administering to the subject an immunogenic amount of an antigenic preparation of the HIV-1_{JC} and optionally (b) harvesting said antibodies to HIV-1_{JC}.

17. The method of claim 16 wherein said subject is a primate.

18. A method of inducing antibodies to HIV-1_{NC} in a nonhuman mammalian subject comprising the steps of (a) administering to the subject an immunogenic amount of an antigenic preparation of the HIV-1_{NC} and optionally (b) harvesting said antibodies to HIV-1_{NC}.

19. The method of claim 18 wherein said subject is a primate.

20. A method of immunizing a subject against the development of acquired immune deficiency syndrome (AIDS) comprising the step of administering to said subject an immunogenic amount of an antigenic preparation of HIV-1_{JC} or HIV-1_{NC} such that antibodies directed to HIV-1_{JC} or HIV-1_{NC} are produced in said subject and thereby symptoms of AIDS are diminished or prevented.

21. The method of claim 20 wherein said subject is a primate.

22. A method for inducing acquired immune deficiency syndrome (AIDS) in a nonhuman primate, comprising the step of administering to said primate an effective amount of an infections preparation of the HIV-1 of claim 1 such that said primate develops AIDS.

23. The method of claim 22 wherein said infections preparation of HIV-1 is a biological specimen obtained from a nonhuman primate infected with HIV-1_{JC} and exhibiting AIDS.

24. The method of claim 22 wherein said infections preparation of HIV-1 is a biological specimen obtained from a nonhuman primate infected with HIV-1_{NC} and exhibiting AIDS.

25. A method of screening for HIV-1 in a biological sample comprising the step of introducing into said biological sample a hybridization probe comprising a nucleotide sequence comprising at least about 15 contiguous nucleotides from SEQ ID NO:11 or a sequence complementary thereto or from SEQ ID NO:12 or a nucleotide sequence complementary thereto under stringent conditions such that said hybridization probe binds to an HIV-1_{JC} gene or a nucleotide sequence having at least about 95% nucleotide sequence identity thereto or an HIV-1_{NC} gene or a nucleotide sequence having at least about 95% nucleotide sequence identity thereto.

26. A nonhuman primate model infected with at least one HIV-1 of claim 1 and exhibiting symptoms of AIDS, useful for the development of a drug or vaccine for the treatment or prevention of AIDS.

27. The primate model of claim 26 wherein said primate is a chimpanzee or a macaque.

28. The primate model of claim 27 wherein said HIV-1 is HIV-1_{JC}.

29. The primate model of claim 27 wherein said HIV-1 is HIV-1_{NC}.

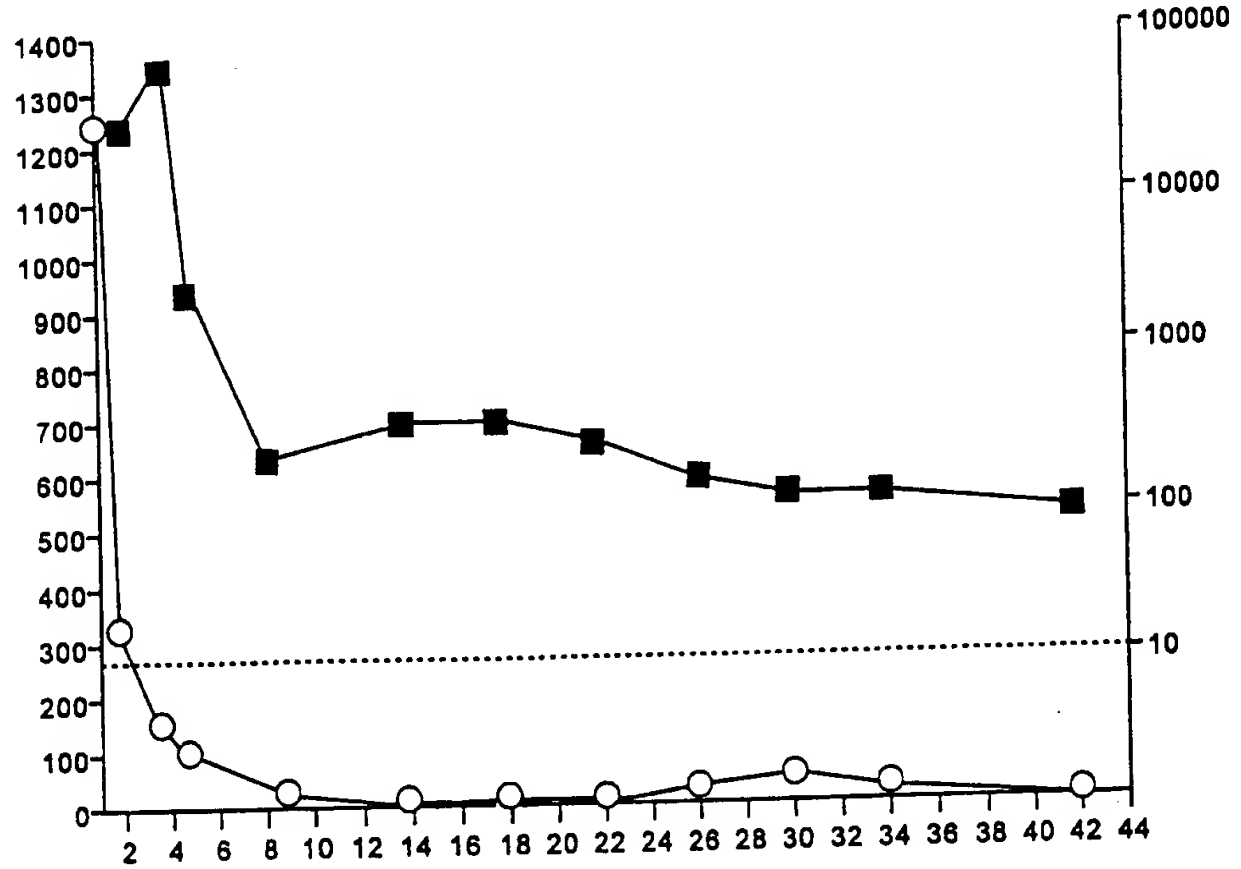
1400
1300
1200
1100
1000
900
800
700
600
500
400
300
200
100
0
100000
10000
1000
100
10
2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36 38 40 42 44
Weeks Post Transfusion
■——■Plasma HIV-1 RNA (Equivalents, 10^3/ml)
○——○Absolute Circulating $CD4^+$ Cells/μl

FIG. 1

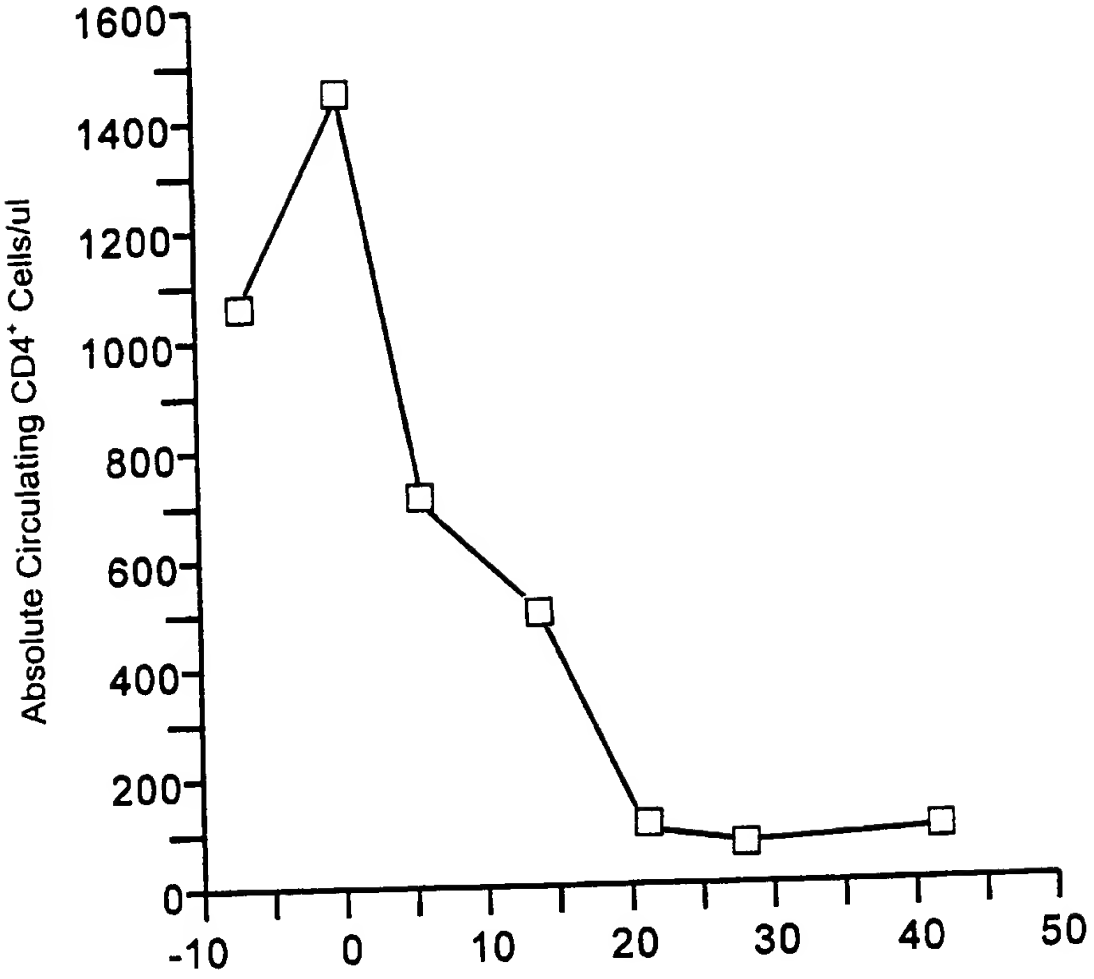
Absolute Circulating CD4+ Cells/ul
1600
1400
1200
1000
800
600
400
200
0
-10
0
10
20
30
40
50
Days Post Infection

FIG. 2

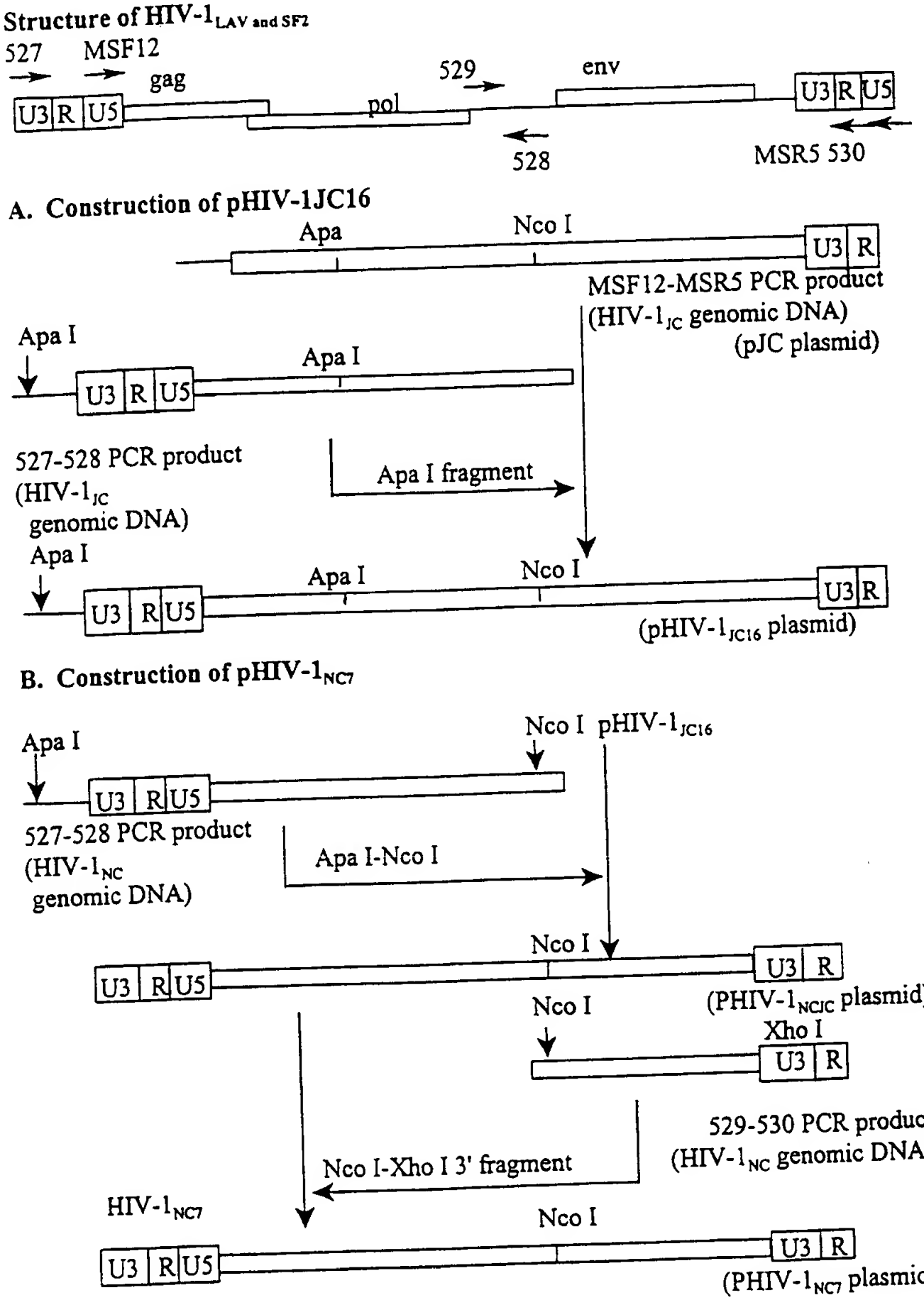
Structure of HIV-1LAV and SF2
527
MSF12
gag
529
env
U3 R U5
pol
528
MSR5 530
A. Construction of pHIV-1JC16
Apa
Nco I
U3 R
MSF12-MSR5 PCR product
(HIV-1JC genomic DNA)
(pJC plasmid)
Apa I
Apa I
U3 R U5
527-528 PCR product
(HIV-1JC
genomic DNA)
Apa I fragment
Apa I
Apa I
Nco I
U3 R U5
U3 R
(pHIV-1JC16 plasmid)
B. Construction of pHIV-1NC7
Apa I
Nco I pHIV-1JC16
U3 R U5
527-528 PCR product
(HIV-1NC
genomic DNA)
Apa I-Nco I
Nco I
U3 R U5
U3 R
(PHIV-1NCJC plasmid)
Nco I
Xho I
U3 R
529-530 PCR product
(HIV-1NC genomic DNA)
Nco I-Xho I 3' fragment
HIV-1NC7
Nco I
U3 R U5
U3 R
(PHIV-1NC7 plasmid)

FIG. 3

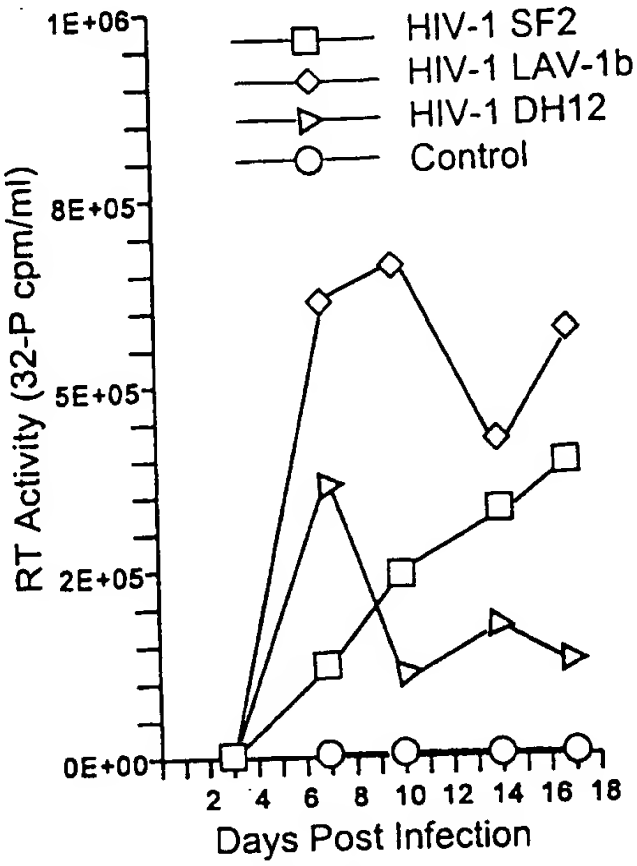

FIG. 4A

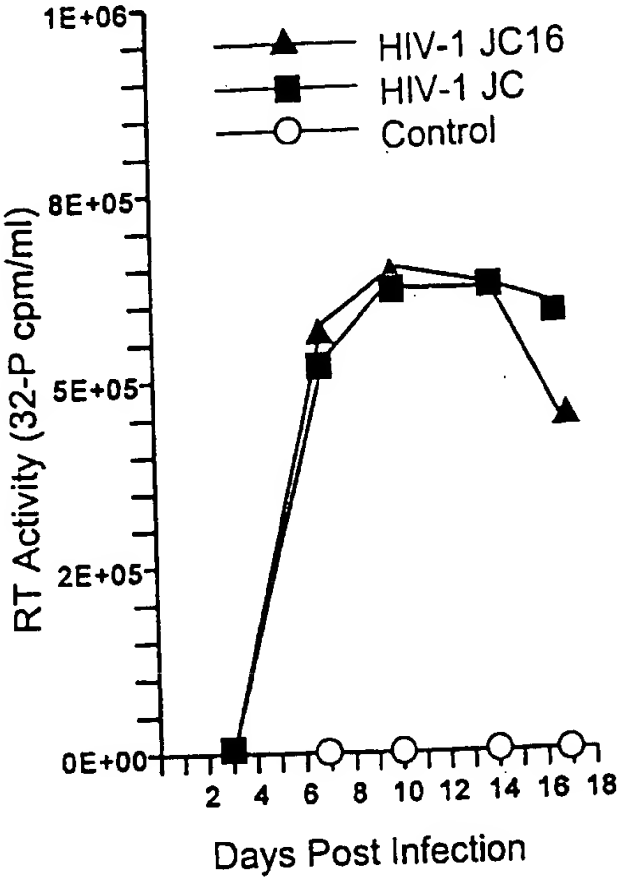

FIG. 4B

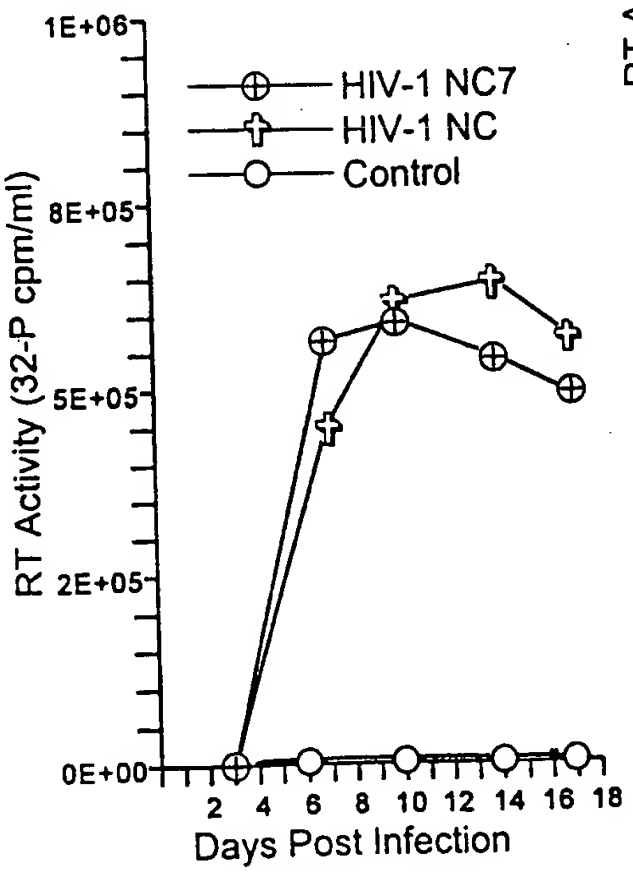

FIG. 4C

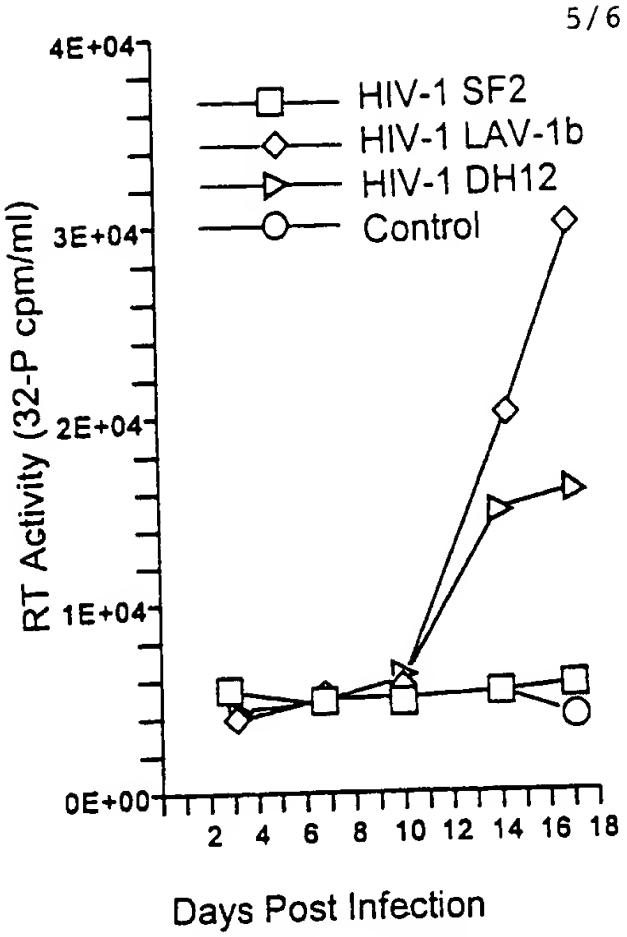

FIG. 4D

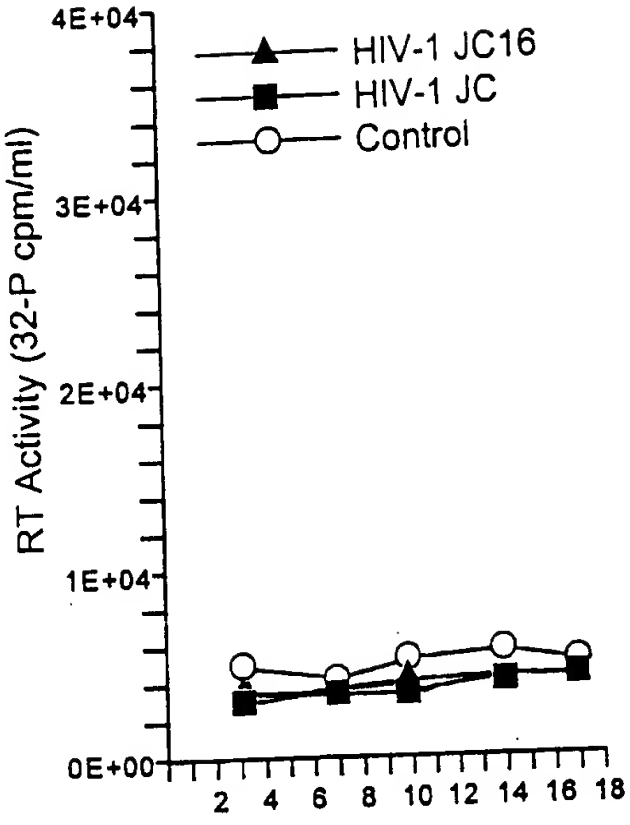

FIG. 4E

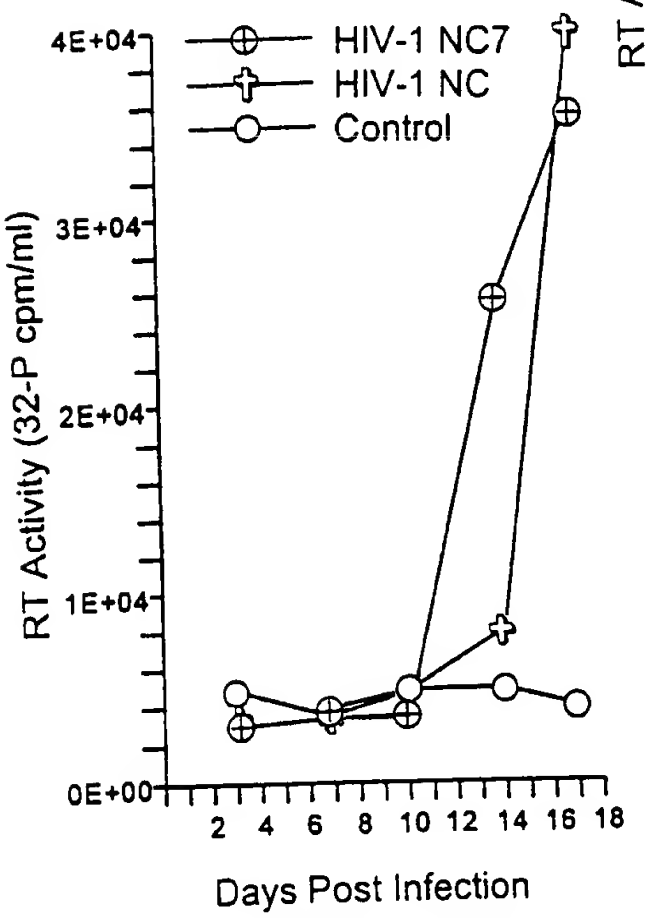

FIG. 4F

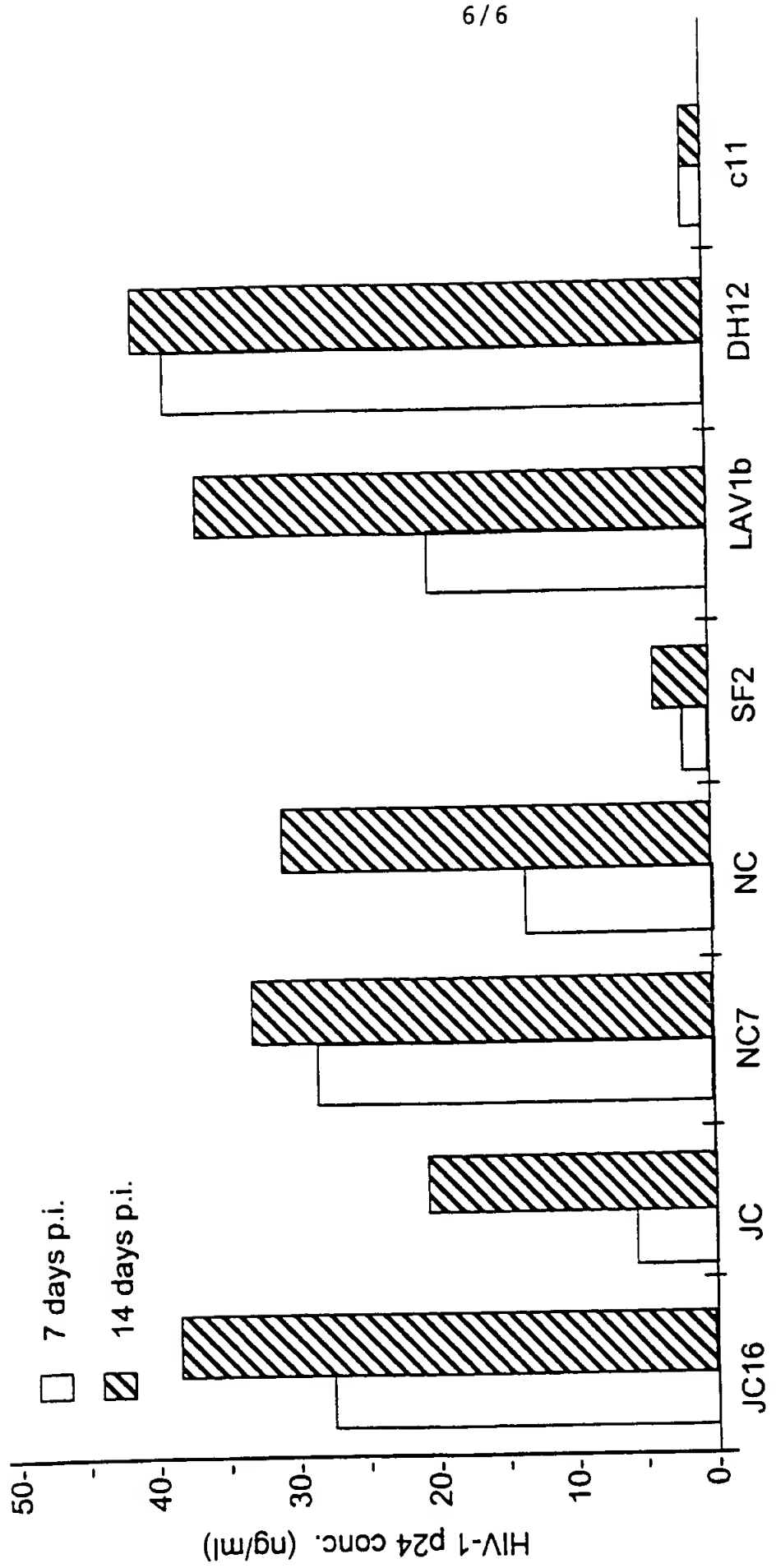
HIV-1 p24 conc. (ng/ml)
50
40
30
20
10
0
7 days p.i.
14 days p.i.
JC16
JC
NC7
NC
SF2
LAV1b
DH12
c11
Virus

FIG. 5

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US98/12990

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :C12Q 1/68, 1/70; C12N 7/00, 7/01, 7/02; G01N 33/564; C07K 1/00

US CL : 435/5, 6, 235, 239, 974; 530/350, 826

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/5, 6, 235, 239, 974; 530/350, 826

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

NONE

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS, Medline, Aidsline

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	NOVEMBRE et al Development of AIDS in a Chimpanzee Infected with Human Immunodeficiency Virus Type 1 Journal Of Virology. May 1997, Vol. 71, No. 5, pages 4086-4091, see entire document.	1-15
X	VINCENT et al. Characterization of a Novel Baboon Virus Closely Resembling Human T-Cell Leukemia Virus Virology. 01 December 1996, Vol. 226, No. 1, pages 57-65, see entire document.	1-15

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search 20 AUGUST 1998	Date of mailing of the international search report 18 SEP 1998
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer JEFFREY STUCKER Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US98/12990

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

Please See Extra Sheet.

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☒ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
1-15 only to the extent of the elected specie, envelope protein.

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US98/12990

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING
This ISA found multiple inventions as follows:

This application contains the following inventions or groups of inventions which are not so linked as to form a single inventive concept under PCT Rule 13.1. In order for all inventions to be searched, the appropriate additional search fees must be paid.

Group I, claims 1-12, drawn to a virus.
Group II, claims 13-15, drawn to viral proteins.
Group III, claims 16-21, drawn to a method of inducing antibodies.
Group IV, claims 22-4 and 26-29, drawn to a primate model.
Group V, claim 25, drawn to a method of screening for HIV infection.
The inventions listed as Groups I-V do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: The rules of PCT unity of invention do not allow for multiple inventions. Each of the inventions requires a different search and a search of each of the inventions would not be co-extensive with each of the other searches.

This application contains claims directed to more than one species of the generic invention. These species are deemed to lack Unity of Invention because they are not so linked as to form a single inventive concept under PCT Rule 13.1. In order for more than one species to be searched, the appropriate additional search fees must be paid. The species are as follows:

Group II: envelope, gag, nef, and p24.

The claims are deemed to correspond to the species listed above in the following manner:

Group II, claims 13-15

The following claims are generic: 1-12.

The species listed above do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, the species lack the same or corresponding special technical features for the following reasons: Each of the proteins is a different compound and has different characteristics such as biological function and immunological reactivity.